



Rui Tiago Fernandes dos Reis

Licenciado em Ciências de Engenharia do Ambiente

**Aplicação de redes neuronais artificiais
na gestão da produção de biogás
resultante de processos de digestão
anaeróbia**

Dissertação para obtenção do Grau de Mestre em Engenharia
do Ambiente, Perfil de Engenharia Sanitária

Orientador: Professora Doutora Leonor Miranda Monteiro
do Amaral, Prof. Auxiliar, FCT-UNL

Co-orientadora: Engenheira Lisete Calado Epifâneo,
Águas de Lisboa e Vale do Tejo, S.A

Júri:

Presidente:	Professor Doutor Pedro Manuel da Hora Santos Coelho
Arguente:	Professor Doutor José Manuel Ribeiro Matos da Fonseca
Vogais:	Professor Doutor António Pedro Macedo Coimbra Mano
	Professora Doutora Leonor Miranda Monteiro do Amaral



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Outubro de 2015



Rui Tiago Fernandes dos Reis

Licenciado em Ciências de Engenharia do Ambiente

**Aplicação de redes neuronais artificiais
na gestão da produção de biogás
resultante de processos de digestão
anaeróbia**

Dissertação para obtenção do Grau de Mestre em Engenharia
do Ambiente, Perfil de Engenharia Sanitária

Orientador: Professora Doutora Leonor Miranda Monteiro
do Amaral, Prof. Auxiliar, FCT-UNL

Co-orientadora: Engenheira Lisete Calado Epifâneo,
Águas de Lisboa e Vale do Tejo, S.A

Júri:

Presidente:	Professor Doutor Pedro Manuel da Hora Santos Coelho
Arguente:	Professor Doutor José Manuel Ribeiro Matos da Fonseca
Vogais:	Professor Doutor António Pedro Macedo Coimbra Mano
	Professora Doutora Leonor Miranda Monteiro do Amaral



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Outubro de 2015

Aplicação de redes neuronais artificiais na gestão da produção de biogás resultante de processos de digestão anaeróbia, “Copyright” Rui Tiago Fernandes dos Reis, da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

“A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor”.

Agradecimentos

Primeiro agradeço a Deus por tudo o que tem feito na minha vida e por me permitir concluir mais esta etapa.

Quero agradecer aos meus pais, por todo o apoio que me deram ao longo do meu percurso escolar e universitário, pois foi absolutamente fundamental. Quero também fazer um agradecimento especial à minha avó paterna por me ter tratado como um filho nestes cinco anos em que vivi na sua casa.

Agradeço à Professora Doutora Leonor Amaral, primeiro por me ter dado a oportunidade de executar este trabalho. Em segundo, por toda a disponibilidade que sempre demonstrou, pelos conhecimentos que me transmitiu e pela motivação que sempre me concedeu.

Quero também agradecer à Eng.^a Lisete Epifânio, pela disponibilidade demonstrada, pelos conhecimentos, informação e diretrizes que me concedeu durante a realização da investigação.

Agradeço ao Eng.^o José Pacheco e à Eng.^a Maria Vieira pela prontidão no envio de dados, pelos conhecimentos que me transmitiram sobre cada ETAR e acima de tudo pela oportunidade de testemunhar o excelente desempenho profissional de ambos.

Agradeço ao Eng.^o Eduardo Ventura, pela disponibilidade que demonstrou e pelo apoio técnico que me concedeu relativamente ao programa de computador NeuralTools[®].

Agradeço ao Grupo Águas de Portugal e à SIMARSUL a disponibilização de todos os dados necessários à realização deste trabalho.

Por último mas não menos importante, quero agradecer aos meus amigos e colegas, por todo o apoio, incentivo e diversão que me concederam durante este período.

Resumo

A investigação realizada integra-se num estudo no âmbito das redes neuronais artificiais (RNA), previamente iniciado em diferentes ETAR do Grupo Águas de Portugal. O trabalho executado centrou-se em dois casos de estudo diferentes, um localizado na ETAR do Seixal e outro situado na ETAR de Sesimbra. O principal alvo da investigação efetuada na ETAR do Seixal foi gerar RNA com capacidade de prever valores de condutividade existentes no caudal total afluente à ETAR, de modo a adicionar este parâmetro a um conjunto de variáveis pré-existente e estudar a sua influência no processo de digestão anaeróbia. Na ETAR de Sesimbra o foco do estudo foi a criação de um conjunto de variáveis afeto ao mesmo processo de tratamento e com base no potencial de aprendizagem das RNA, estimar valores de biogás gerado. Concluída a investigação em cada ETAR, criou-se um modelo comum aos dois casos de estudo com o objetivo de prever valores de biogás gerado.

As RNA foram produzidas com recurso a um programa de computador denominado NeuralTools[®]. As séries de dados das diferentes variáveis constituíram a base de informação das RNA e a sua construção e preparação foi um aspeto fundamental no trabalho realizado. A análise estatística foi o principal suporte do tratamento de dados efetuado em ambos os casos de estudo. Foram produzidas bastantes RNA e em diferentes condições, de forma a gerar os melhores resultados possíveis.

Na investigação concretizada na ETAR do Seixal não se produziram bons resultados, por insuficiência do conhecimento da relação entre o funcionamento do processo de digestão anaeróbia e os valores de condutividade registados. Na ETAR de Sesimbra, a melhor RNA foi produzida com base nas séries de dados que evidenciaram um maior coeficiente de correlação com a série de dados de caudal de biogás. Os resultados gerados na ETAR de Sesimbra foram satisfatórios, tal como também foram os resultados dos testes realizados com o modelo global. O principal obstáculo na produção de melhores resultados advém da elevada variabilidade de valores existente nas séries de dados relacionadas com o funcionamento do processo de digestão anaeróbia.

Palavras-chave: Digestão anaeróbia, Redes neuronais artificiais, Biogás, Condutividade, Previsão.

Abstract

The research focuses on the study of artificial neural networks (ANN), previously initiated in different WWTP that belong to the “Águas de Portugal” Group. The work performed focused on two different study cases, one located at the WWTP of Seixal and another located in WWTP of Sesimbra. The main target of the research being conducted at the WWTP Seixal was to generate ANN capable of providing conductivity values on total flow tributary to the WWTP, add this parameter to a set of pre-existing variables and study its influence on the anaerobic digestion process. In WWTP of Sesimbra the study focus was the creation of a set of variables linked to the same treatment process and based on the ANN learning potential and to predict the biogas generated values. In the end of the investigation in each WWTP, it was created a common model to the two study cases in order to predict biogas generated values.

The ANN were built using the NeuralTools[®] software. The different variables data series were the ANN information database and its construction and preparation was the key feature of this paper. Statistical analysis was the backbone process performed in both study cases. Sufficient ANN were produced under different conditions in order to generate the best possible results.

The investigation overseen at the WWTP of Seixal did not produce good results, primarily due to insufficient knowledge of the relationship between the operation of the anaerobic digestion process and the registered conductivity values. At the WWTP of Sesimbra the best ANN was produced based on the data series that showed a higher correlation coefficient with the biogas flow data series. The results generated at this WWTP were satisfactory, as were also the results of the tests performed with the common model. When trying to produce the most reliable results, the main obstacle is certainly the high variability of values in the data series related to the operation of the anaerobic digestion process.

Keywords: Anaerobic digestion, Artificial neural networks, Biogas, Conductivity, Forecast.

Índice de Matérias

1. Introdução	1
1.1 Problemática	1
1.2 Enquadramento da Dissertação	3
1.3 Objetivos	3
1.4 Organização da Dissertação	5
2. Revisão Bibliográfica	7
2.1 Águas Residuais	7
2.2 Tratamento de Águas Residuais	8
2.2.1 Legislação em Portugal	8
2.2.2 Estações de Tratamento de Águas Residuais	11
2.3 Processo de Digestão Anaeróbia	15
2.3.1 Vantagens e Desvantagens da Digestão Anaeróbia	16
2.3.2 Teoria do Processo de Digestão Anaeróbia	17
2.3.3 Tipos de Digestão Anaeróbia	20
2.3.4 Dimensionamento do Sistema de Digestão Anaeróbia	23
2.3.5 Variáveis ambientais que Influenciam o Processo de Digestão Anaeróbia	25
2.3.6 Influência da Salinidade no Funcionamento da Digestão Anaeróbia	28
2.3.7 Produtos da Digestão Anaeróbia	33
2.3.8 Modelação: Principais Tipologias Aplicadas	34
2.4 Redes Neurais Artificiais	37
2.4.1 Funcionamento do Neurónio Celular	37
2.4.2 Evolução histórica das Redes Neurais Artificiais	38
2.4.3 Funcionamento das Redes Neurais Artificiais	39
2.4.4 Tipos de Redes Neurais Artificiais	42
3. Metodologia	49
3.1 Descrição Sumária	49
3.2 Entidades Envolvidas nos Casos de Estudo	49
3.2.1 Grupo Águas de Portugal	49
3.2.2 SIMARSUL	51
3.3 Casos de Estudo	52

3.3.1 Caso de Estudo - ETAR do Seixal	53
3.3.2 Caso de Estudo - ETAR de Sesimbra	62
4. NeuralTools®	71
4.1 Funcionamento do NeuralTools®	71
5. Resultados e Discussão	83
5.1 Caso de Estudo – ETAR do Seixal.....	83
5.1.1 Recolha de Dados e Construção das Novas Séries de Dados	83
5.1.2 Previsão de Valores de Produção de Biogás	89
5.1.3 Análise da Influência de Intrusões Salinas no Processo de Digestão Anaeróbia	95
5.2 Caso de Estudo – ETAR de Sesimbra	107
5.2.1 Recolha e Preparação de Dados	107
5.2.2 Etapa de Treino e Teste.....	117
5.2.3 Etapa de Previsão.....	128
5.3 Modelo Geral Comum aos Dois Casos de Estudo	133
6. Conclusão e Sugestões para Desenvolvimentos Futuros	139
7. Referências Bibliográficas.....	143
Anexo I	153
Anexo II	157
Anexo III	159
Anexo IV	163
Anexo V	165
Anexo VI	167
Anexo VII	169
Anexo VIII	171

Índice de Figuras

Figura 2.1 – Diagrama simplificado das reações presente no processo de digestão anaeróbia	18
Figura 2.2 – Digestão anaeróbia em baixa carga	21
Figura 2.3 – Digestão anaeróbia em alta carga	22
Figura 2.4 – Procedimento esquemático dos modelos dinâmicos.....	35
Figura 2.5 – Estrutura simplificada do neurónio celular	37
Figura 2.6 – Fluxo de informação na biologia celular face à representação esquemática desse fluxo em um modelo matemático.....	38
Figura 2.7 – Modelo de um neurónio artificial	39
Figura 2.8 – Representação esquemática típica da arquitetura de uma rede de multicamada feedforward	44
Figura 2.9 – Representação esquemática da arquitetura de uma rede recorrente de uma só camada.....	45
Figura 2.10 – Representação esquemática da arquitetura de rede probabilística	46
Figura 2.11 – Representação esquemática da arquitetura de rede de função de base radial.....	47
Figura 3.1 – Empresas de abastecimento de água e saneamento de águas residuais	50
Figura 3.2 – Municípios abrangidos pela SIMARSUL	51
Figura 3.3 – Concelho do Seixal e local de Implantação e vista aérea da ETAR.....	53
Figura 3.4 – Planta do sistema de tratamento da ETAR do Seixal.....	54
Figura 3.5 – Metodologia aplicada na investigação realizada na ETAR do Seixal	59
Figura 3.6 – Concelho de Sesimbra e local de implementação e vista aérea da ETAR	62
Figura 3.7 – Planta do sistema de tratamento da ETAR de Sesimbra	63
Figura 4.1 – Localização da barra de ferramentas do NeuralTools® na interface do Microsoft Excel®.	72
Figura 4.2 – Barra de ferramentas do NeuralTools®, onde estão inseridos os comandos de controlo do programa.	72
Figura 4.3 – Estrutura de dados, formada por variáveis e casos.....	73
Figura 4.4 – Caixa de diálogo do comando Data Set Manager	73
Figura 4.5 – Caixa de diálogo do comando Train	75
Figura 4.6 – Secção referente à configuração de rede do comando Train.....	76
Figura 4.7 – Secção referente às condições de paragem de treino, do comando Train	77

Figura 4.8 – Caixa de diálogo onde são exibidas as condições em que o treino decorrerá....	78
Figura 4.9 – Caixa de diálogo, onde é possível acompanhar o processo de treino das redes neuronais.....	78
Figura 4.10 – Caixa de diálogo do comando Predict	80
Figura 4.11 – Caixa de diálogo com o resumo das definições da etapa de previsão.....	81
Figura 4.12 – Subcomandos do comando Utilities.....	82
Figura 5.1 – Distribuição temporal da variável “Q biogás” – ETAR do Seixal	85
Figura 5.2 – Frequência de valores registados da série de dados da variável “Q biogás” – ETAR do Seixal	85
Figura 5.3 – Distribuição temporal das variáveis “Q biogás” e “Q lamas” – ETAR do Seixal.....	86
Figura 5.4 – Distribuição temporal das variáveis “Q biogás” e “T reator” – ETAR do Seixal...	86
Figura 5.5 – Dados de condutividade e da altura de maré de 16 de Maio a 21 de Maio de 2015.....	97
Figura 5.6 - Dados de condutividade e de caudal médio horário de 16 de Maio a 21 de Maio de 2015.....	97
Figura 5.7 – Metodologia aplicada na etapa de treino e teste das redes neuronais criadas com o intuito de prever valores de condutividade na ETAR do Seixal	99
Figura 5.8 – Média horária dos valores de “Q instantâneo”	104
Figura 5.9 – Evolução da “T reator” – ETAR de Sesimbra.	113
Figura 5.10 – Distribuição temporal da variável “Q efluente bruto” – ETAR de Sesimbra.....	113
Figura 5.11 – Distribuição temporal das variáveis “Q lamas” – ETAR de Sesimbra	114
Figura 5.12 – Distribuição temporal da variável “Q biogás” – ETAR de Sesimbra	115
Figura 5.13 – Metodologia aplicada na etapa de treino e teste das redes neuronais criadas com o intuito de prever valores de biogás produzido na ETAR de Sesimbra ...	117
Figura 5.14 – Média dos resultados gerados por número de dias de desfasamento	121
Figura 5.15 – Metodologia aplicada na etapa de previsão na ETAR de Sesimbra	128
Figura 5.16 – Comparação entre os valores reais e os valores previstos de biogás produzido pelas redes “1 9G8 sx” e “3I” – ETAR de Sesimbra	130
Figura 5.17 – Comparação entre os valores reais e os valores previstos de biogás produzido pelas redes “2E” e “4E” – ETAR de Sesimbra	132
Figura 5.18 – Comparação dos valores reais e dos valores previstos de biogás produzido pelas redes “Geral sx” e “Geral sb” – ETAR de Sesimbra.....	135
Figura 5.19 – Diagrama da investigação realizada	137

Figura I.1 –	Relação entre os valores reais, previstos e residuais.....	154
Figura II.1 –	Localização das medições efetuadas e locais de amostragem de alguns parâmetros afetos à digestão anaeróbia – ETAR do Seixal.....	157
Figura II.2 –	Local onde é efetuada a medição afeta ao caudal de biogás – ETAR do Seixal.....	157
Figura III.1 –	Comparação entre os dados das variáveis “pH in” e “pH reator” – ETAR do Seixal.....	159
Figura III.2 –	Comparação entre os dados das variáveis “CQO in” e “CQO reator” – ETAR do Seixal.....	159
Figura III.3 -	Comparação entre os dados das variáveis “LCQO in” e “LCQO reator” – ETAR do Seixal	160
Figura III.4 -	Comparação entre os dados das variáveis “MV in” e “MV reator” – ETAR do Seixal.....	160
Figura III.5 -	Comparação entre os dados das variáveis “LMV in” e “LMV out” – ETAR do Seixal.....	160
Figura III.6 -	Comparação entre os dados das variáveis “TAL reator” e “AGV reator” – ETAR do Seixal	161
Figura VI.1 –	Localização das medições efetuadas e locais de amostragem de alguns parâmetros afetos à digestão anaeróbia – ETAR de Sesimbra	167
Figura VI.2 -	Local onde é efetuada a medição afeta ao caudal de biogás – ETAR de Sesimbra	167
Figura VII.1 –	Comparação entre os dados das variáveis “pH in” e “pH reator” – ETAR de Sesimbra	169
Figura VII.2 –	Comparação entre os dados das variáveis “MV in” e “MV reator” – ETAR de Sesimbra	169
Figura VII.3 –	Comparação entre os dados das variáveis “LMV in” e “LMV reator” – ETAR de Sesimbra	169
Figura VII.4 –	Comparação entre os dados das variáveis “TAL reator” e “AGV reator” – ETAR de Sesimbra.....	170

Índice de Tabelas

Tabela 2.1 – Principais propriedades físicas, químicas e biológicas que compõem as águas residuais	7
Tabela 2.2 – Requisitos para as descargas das estações de tratamento de águas residuais	9
Tabela 2.3 – Requisitos para as descargas de águas residuais urbanas em zonas sensíveis sujeitas a eutrofização	9
Tabela 2.4 – Níveis de tratamento de águas residuais	12
Tabela 2.5 – Critérios de dimensionamento para digestores anaeróbios	23
Tabela 2.6 – Relação entre a temperatura de operação e o tempo de retenção de sólidos em digestores anaeróbios de alta carga	26
Tabela 2.7 – Concentrações inibitórias de compostos presentes em processos anaeróbios	28
Tabela 2.8 – Concentração dos iões principais na superfície da água do mar	29
Tabela 3.1 – Dados de atividade do Grupo Águas de Portugal em 2014	50
Tabela 3.2 – Dados do sistema em exploração	52
Tabela 3.3 – Dados do sistema relativos ao ano horizonte de projeto	52
Tabela 3.4 – Dados de base da ETAR do Seixal	54
Tabela 3.5 – Dados de base da ETAR de Sesimbra	63
Tabela 5.1 – Variáveis transpostas para a presente investigação	83
Tabela 5.2 – Número de dados disponíveis para a produção de RNA	84
Tabela 5.3 – Média dos valores registados afetos às variáveis “Q lamas”, “T reator” e “Q biogás” – ETAR do Seixal	86
Tabela 5.4 – Resultados do cálculo do coeficiente de correlação de Pearson das séries de dados afetas ao digestor anaeróbio da ETAR do Seixal	88
Tabela 5.5 – Cinco melhores redes neuronais produzidas no trabalho de 2014	90
Tabela 5.6 – Resultados da fase de previsão por rede neuronal	90
Tabela 5.7 – Resultados globais da repetição da produção das cinco melhores redes neuronais de 2014	92
Tabela 5.8 – Resultados da fase de previsão da repetição da produção das cinco melhores redes neuronais de 2014	92
Tabela 5.9 – Resultados quando destacados casos para previsão, anteriormente utilizados nas fases de treino e teste de uma rede neuronal	94

Tabela 5.10 – Valores médios de condutividade, altura de maré e caudal instantâneo, por período temporal	98
Tabela 5.11 – Resultados gerados pelas redes neuronais de classe B	100
Tabela 5.12 – Resultados gerados pelas redes neuronais de classe C	101
Tabela 5.13 – Resultados das redes neuronais de classe B, referentes ao segundo estágio .	102
Tabela 5.14 – Resultados gerados pela rede neuronal “C4”	102
Tabela 5.15 – Resultados gerados pela rede neuronal “L1”	103
Tabela 5.16 – Resultados gerados pelas redes neuronais “L2” e “L3”	104
Tabela 5.17 – Resultados das redes neuronais geradas no quarto estágio	106
Tabela 5.18 – Variáveis consideradas para integrar o estudo	108
Tabela 5.19 – Resultados do cálculo do coeficiente de correlação de Pearson das séries de dados afetas ao segundo caso de estudo	110
Tabela 5.20 – Médias mensais de alguns parâmetros associados ao processo de digestão aneróbia	112
Tabela 5.21 – Estatísticas das séries de dados de frequência de amostragem diária.....	112
Tabela 5.22 – Distribuição de dados da variável “Q lamas” por intervalo de valores.....	114
Tabela 5.23 – Estatísticas da série de dados da variável “TRH” utilizadas nos testes com as RNA	114
Tabela 5.24 – Distribuição de dados da variável “Q biogás” por intervalo de valores.....	115
Tabela 5.25 – Estatísticas das séries de dados de frequência de amostragem semanal.....	116
Tabela 5.26 – Valores de referência para os parâmetros de frequência de amostragem semanal.....	116
Tabela 5.27 – Resultados das redes neuronais “1A1”, “2A1” e “3A1”	119
Tabela 5.28 – Resultados dos novos testes realizados com as configurações das redes “1A1”, “2A1” e “3A1”	119
Tabela 5.29 – Média dos resultados gerados por número de dias de desfasamento	121
Tabela 5.30 – Número de utilizações de cada variável na produção das redes neuronais	122
Tabela 5.31 - Resultados das melhores redes neuronais geradas a partir de séries de dados de frequência de amostragem semanal, sem desfasamento de dados .	124
Tabela 5.32 – Número de presenças de cada variável nas configurações das melhores redes neuronais criadas	125
Tabela 5.33 – Resultado da nova melhor rede “1 9G8 sx”	126
Tabela 5.34 – Frequência de intervalos entre amostragens.....	126

Tabela 5.35 – Melhores redes neuronais de séries de dados de frequência de amostragem semanal, com desfasamento entre dados	127
Tabela 5.36 – Melhor resultado na fase de previsão das redes neuronais com desfasamento de dados	129
Tabela 5.37 – Melhores resultados na fase de previsão das redes neuronais com dados de frequência de amostragem diária.....	129
Tabela 5.38 – Resultados da fase de previsão das redes neuronais de dados de frequência de amostragem semanal, sem desfasamento	130
Tabela 5.39 – Resultados da nova fase de previsão para as redes “2E” e “4E”	131
Tabela 5.40 – Relação entre as correlações mais fortes verificadas nos dois casos de estudo.....	133
Tabela 5.41 – Relação entre as correlações mais fortes verificadas com a variável “Q biogás”	134
Tabela 5.42 – Resultados das fases de desenvolvimento das redes neuronais “Geral sx” e “Geral sb”	135
Tabela I.1 – Exemplo de um relatório detalhado gerado após o término da fase de treino e teste.....	153
Tabela IV.1 – Resultados do cálculo do coeficiente de correlação de Pearson	163
Tabela V.1 – Casos considerados na etapa de previsão das redes neuronais geradas a partir das novas séries de dados existentes na ETAR do Seixal	165
Tabela VIII.1 –Resultados da fase de treino e teste das redes “2 8H sx” e “3I”	171

Lista de Abreviaturas, Acrónimos, Siglas e Símbolos

Abreviaturas

e.p. – Equivalente de população

Neural AD – Neural Networks + Anaerobic Digestion

Acrónimos

AdP – Águas de Portugal, SGPS, S.A.

DQA – Diretiva Quadro da Água

EGF - Empresa Geral do Fomento

ETAR – Estação de Tratamento de Águas Residuais

FCT/UNL – Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

IWA – International Water Association

PEAASAR – Plano Estratégico de Abastecimento de Água e Saneamento de Águas Residuais

PENSAAR – Plano Estratégico Nacional para o Setor do Abastecimento de Águas e Saneamento de Águas Residuais

QREN – Quadro de Referência Estratégica Nacional

RNA – Redes Neurais Artificiais

SIMARSUL – Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal, S.A.

SIMRIA – Sistema Multimunicipal de Saneamento da Ria de Aveiro, S.A.

UE – União Europeia

Siglas

PM – Preia-mar

BM – Baixa-mar

Simbologia

C – Carbono

Ca - Cálcio

CH₃COOH– Ácido acético

CH₄ – Metano

Cl - Cloro

CO₂ – Dióxido de carbono

H₂ – Hidrogénio

H₂O – Água

K - Potássio

Mg – Magnésio

N – Azoto

Na - Sódio

NO₃ – Azoto de nitratos

NO₂ – Azoto de nitritos

NaCl – Cloreto de sódio

O₂ – Oxigénio

P – Fósforo

°C – Grau Celsius

atm - atmosfera

g – Grama

h – Hora

J - Joule

L – Litro

m – Metro

Nm³ – Metro cúbico normal (medido em condições standard de pressão e temperatura)

S - Siemens

AOV - Ácidos orgânicos voláteis

AGV – Ácidos gordos voláteis

AGV reator – Concentração de ácidos gordos voláteis nas lamas digeridas (mg CH₃COOH/L)

AGV/TAL – Quociente entre a concentração de AGV no digestor e a alcalinidade total

ATP – Trifosfato de Adenosina

CBO₅ – Carência bioquímica de oxigénio ao fim de cinco dias (mg/L)

COT – Carbono orgânico total (mg/L)

CQO – Carência química de oxigénio (mg/L)

CQO in – CQO medida nas lamas afluentes ao digestor anaeróbio (mg/L)

CQO reator - CQO medida nas lamas digeridas (mg/L)

CTO – Carência total de oxigénio (mg/L)

CV – Carga orgânica volúmica ($\text{kg}/(\text{m}^3 \cdot \text{dia})$)

LCQO in – Carga de CQO afluente ao digestor anaeróbio (kg/dia)

LCQO out - Carga de CQO nas lamas digeridas (kg/dia)

LCQO removida – Carga de CQO removida na digestão anaeróbia (kg/dia)

LMS in - Carga de matéria seca afluente à digestão anaeróbia (kg/dia)

LMV in - Carga de matéria volátil afluente à digestão anaeróbia (kg/dia)

LMV out - Carga de matéria volátil nas lamas digeridas (kg/dia)

LMV removida – Carga de matéria volátil removida na digestão anaeróbia (kg/dia)

MS in – Concentração de sólidos totais nas lamas afluentes à digestão anaeróbia (mg/L)

MS reator - Concentração de sólidos totais nas lamas digeridas (mg/L)

MV in - Concentração de sólidos voláteis nas lamas afluentes à digestão anaeróbia (mg/L)

MV reator - Concentração de sólidos voláteis nas lamas digeridas (mg/L)

pH – potencial de hidrogénio

pH in – pH das lamas afluentes ao digestor anaeróbio

pH reator – pH das lamas digeridas

Q biogás – Caudal de biogás produzido no digestor anaeróbio (m^3/dia)

Q efluente bruto – Caudal de água bruta que entra na estação de tratamento de águas residuais

Q lamas - Caudal de lamas elevadas para o digestor anaeróbio (m^3/dia)

SST – Sólidos suspensos totais (mg/L)

T reator – Temperatura no digestor anaeróbio ($^{\circ}\text{C}$)

TAL reator - Alcalinidade total nas lamas digeridas ($\text{mg CaCO}_3/\text{L}$)

TRH – Tempo de retenção hidráulico no digestor anaeróbio (dias)

UV – Ultravioleta

ADM1 – Anaerobic Digestion Model No. 1

GRNN - Generalized Regression Neural Network

MLFN - Multilayer Feed-forward Network

PNN – Probabilistic Neural Network

RMS – Erro Médio Quadrático

R^2 – Coeficiente de Determinação

r – Coeficiente de correlação de Pearson

1. Introdução

1.1 Problemática

Desde finais do século XX, a preocupação pelas questões ambientais e o impacto destas na economia, a nível europeu e mundial têm aumentando de forma significativa. Nos últimos anos o elevado aumento do preço dos combustíveis fósseis, nomeadamente do petróleo e as preocupações acrescidas com as alterações climáticas levaram as instituições financeiras a uma reavaliação do setor energético e à sua política. A patente necessidade de reduzir os níveis de poluição bem como os custos energéticos, levou a que fosse necessário procurar energias alternativas aos combustíveis fósseis (Jäger-Waldau *et al*, 2011).

As políticas e acordos mais recentes realizados na União Europeia (UE) definiram o objetivo de garantir no ano de 2020, que 20% das necessidades energéticas europeias sejam satisfeitas por sistemas de energias renováveis. Uma parte considerável da energia renovável a produzir terá como fonte de alimentação a biomassa, estimando-se que 25% de toda a bioenergia gerada será utilizada sob a forma de biogás (Holm-Nielsen *et al.*, 2009). Este gás de origem biológica é composto por 65% a 70% de metano (CH_4) e 25% a 30% de dióxido de carbono (CO_2), possuindo ainda pequenas quantidades de outros gases (Metcalf & Eddy, 2003). Têm sido desenvolvidas várias tecnologias com capacidade para converter biomassa em bioenergia, sendo que uma das que mais se destacam é a digestão anaeróbia (Jingura e Matengaifa, 2007). Esta tecnologia além de conseguir gerar biogás a partir da matéria orgânica, possui a vantagem de o substrato digerido, poder ser muitas vezes utilizado na agricultura como fertilizante (Holm-Nielsen *et al.*, 2009).

O aumento de interesse pela produção de biogás como fonte de energia, também é uma realidade na área de Engenharia Sanitária, nomeadamente nas estações de tratamento de águas residuais (ETAR) (Metcalf & Eddy, 2003). Nos últimos anos, a preocupação com a vulnerabilidade do meio ambiente, tem sido uma constante neste ramo. Existe hoje, o dever de melhorar cada vez mais a qualidade do efluente tratado, ao minimizar a matéria orgânica e os nutrientes presentes na altura da descarga do efluente, o que se atinge com custos operacionais bastante elevados associados ao aumento de eficiência dos equipamentos e a utilização de tecnologias sofisticadas (Puig *et al.*, 2011). Quaisquer oportunidades de atenuar estes custos são importantes para as empresas ligadas ao saneamento, pelo que, possuir uma tecnologia no sistema de tratamento, com capacidade de produzir energia para uso na instalação é extremamente vantajoso (Metcalf & Eddy, 2003). No atual panorama, uma boa operação e manutenção de uma infraestrutura deste tipo constitui um desafio cada vez maior a nível económico e ambiental (Puig *et al.*, 2011).

A digestão anaeróbia é um dos processos mais antigos utilizados em ETAR para a estabilização de sólidos e lamas, devido precisamente aos benefícios que proporciona. Em vários casos, os digestores anaeróbios, além da importância que detêm no sistema de tratamento, possuem a capacidade de gerar biogás em quantidade suficiente para colmatar grande parte das necessidades energéticas das estações (Metcalf & Eddy, 2003). Apesar de a digestão anaeróbia conter muitas vantagens, o seu funcionamento pode ser bastante complexo e por vezes complicado de gerir. Este processo possui vários parâmetros que devem ser judiciosamente geridos e é algo sensível à presença de tóxicos, os quais podem, em casos extremos, inibir totalmente a sua ação dos microrganismos, ou no mínimo afetar a eficiência do tratamento e com a consequente diminuição da produção de biogás (Lin Chen *et al.*, 2014).

O facto de ser essencial uma estação de tratamento trabalhar sempre muito perto da sua eficiência máxima, e as características e qualidade do efluente puderem mudar rapidamente, tem levado a que o número de estudos na área de tratamento de águas residuais tem vindo a aumentar. Face aos novos desafios existentes no ramo da Engenharia Sanitária, as empresas têm procurado desenvolver novos métodos de monitorização e controlo, que permitam gerir melhor cada etapa de tratamento. Desde há algum tempo que se usam modelos dinâmicos, criados por tecnologia computacional, capazes de otimizar a performance dos equipamentos, através da seguinte forma (Metcalf & Eddy, 2003):

- Simulam cada processo em diferentes condições de operação;
- Analisam a evolução do funcionamento de cada órgão, o que permite antecipar problemas;
- Estes modelos podem ser integrados no sistema de controlo dos equipamentos.

Estes modelos matemáticos computacionais têm permitido bastantes avanços e melhorias no tratamento de águas residuais (Metcalf & Eddy, 2003). No entanto o processo da digestão anaeróbia ainda é complicado de gerir. Depois de várias tentativas para se conhecer e controlar melhor o funcionamento dos digestores anaeróbios, é um facto que de uma forma geral os digestores ainda são vistos como uma espécie de “caixa negra”, visto que alguns mecanismos do processo ainda não foram totalmente compreendidos. Não se dominando totalmente o funcionamento do digestor, torna-se difícil criar um modelo dinâmico eficiente, dado este ser assente em algumas simplificações processuais, que neste caso não são geradas com exatidão (Holubar *et al.*, 2002).

Nos últimos anos como alternativa à utilização dos modelos mais clássicos, tem-se vindo a introduzir novas ferramentas como as Redes Neurais Artificiais (RNA), as quais têm demonstrado algum potencial para otimizar o funcionamento dos sistemas de tratamento (Holubar *et al.*, 2002). As RNA surgiram pela primeira vez na década de quarenta, mas só começaram a ser amplamente usadas em diversas áreas a partir da década de oitenta (Ding *et al.*, 2011). Esta ferramenta, relativamente a outros métodos, apresenta a vantagem de não requerer o conhecimento prévio das relações entre os diferentes parâmetros e variáveis afetos

a um processo. A vantagem referida pode tornar as RNA numa alternativa bastante interessante, para aplicar na modelação de processos de digestão anaeróbia (Holubar *et al.*, 2002). Esperando-se que à medida que se aplica a novos casos, possa ser possível desbravar o caminho para novos desenvolvimentos na modelação e gestão do funcionamento de um digestor anaeróbio.

1.2 Enquadramento da Dissertação

A presente dissertação enquadra-se num projeto desenvolvido pelo Grupo Águas de Portugal (AdP). O projeto denominado Neural AD (*Neural Networks + Anaerobic Digestion*) visa otimizar a produção de energia a partir do biogás resultante do processo de digestão anaeróbia das lamas geradas nas ETAR, utilizando as RNA como ferramenta de apoio.

A empresa AdP em colaboração com o Instituto Superior de Engenharia do Porto iniciou o Neural AD no ano de 2013, tendo como caso de estudo o digestor anaeróbio da ETAR Sul do Sistema Multimunicipal de Saneamento da Ria de Aveiro (SIMRIA). No final deste estudo, concluiu-se que as RNA demonstraram um potencial promissor na modelação da etapa de digestão anaeróbia. Os resultados verificados motivaram um segundo ano (2014) de investigação no projeto, o qual foi aberto a novas entidades. Com a entrada no projeto do Instituto Superior Técnico, da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL) e da Universidade do Minho, foi possível estudarem-se novos casos e ampliar o campo de aplicações.

A FCT/UNL em consonância com o Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal (SIMARSUL), no ano de 2014, aplicou este projeto a um digestor anaeróbio da ETAR do Seixal. Os resultados apesar de auspiciosos foram limitados pela extensão de dados disponível não ser suficientemente grande. Na atual dissertação existe a responsabilidade de dar continuação ao desenvolvimento do estudo iniciado por estas duas entidades no ano de 2014.

1.3 Objetivos

A dissertação tem como objetivo principal analisar dois casos de estudo diferentes, no âmbito da produção de energia a partir do biogás produzido durante o processo de digestão anaeróbia das lamas geradas nas ETAR, recorrendo às RNA como ferramenta de apoio. O primeiro caso de estudo diz respeito à etapa de digestão anaeróbia na ETAR do Seixal, enquanto o segundo caso de estudo diz respeito ao mesmo processo, mas neste caso na ETAR de Sesimbra. O programa de computador usado para produção das RNA durante este trabalho, designa-se de NeuralTools[®] e foi desenvolvido pela Palisade[®].

Na ETAR do Seixal, o estudo apresenta duas vertentes, a primeira visa continuar a investigação iniciada no mesmo local no ano de 2014, tendo como principais metas a cumprir as seguintes:

- Preparar novas séries de dados, tendo por base o conjunto de dados criado no anterior trabalho;
- Realizar a previsão de valores de produção de biogás gerado durante a digestão anaeróbia, analisando e comparando em seguida os resultados verificados face aos do ano de 2014.

A segunda vertente do trabalho na ETAR do Seixal tem em vista a análise da influência de ingerências salinas na eficiência do processo de digestão anaeróbia. Na rede de drenagem que conduz à ETAR do Seixal verificam-se algumas intrusões de água salina, dadas as condutas serem antigas, apresentando por vezes condições deficitárias e estarem em contacto direto com o Estuário do Rio Tejo em alguns locais. Os valores mais elevados de condutividade registados na ETAR estão em consonância, com os valores mais elevados de altura de maré registados no Rio Tejo. Neste campo, pretende-se atingir os seguintes objetivos:

- Recolher e preparar um conjunto de dados de condutividade;
- Recolher dados de altura de maré e executar curvas de maré;
- Preparar séries de dados e realizar a previsão dos valores de condutividade no caudal afluente à ETAR;
- Estimar os valores de condutividade média diária e utilizar a condutividade como uma das variáveis a ter em conta, na previsão de valores de produção de biogás gerado durante a digestão anaeróbia e efetuar uma análise aos resultados;
- Avaliar o potencial demonstrado pelas RNA na modelação do processo de digestão anaeróbia.

Na ETAR de Sesimbra localiza-se o segundo caso de estudo do atual trabalho, onde se espera concretizar os seguintes objetivos:

- Criar um conjunto de dados a partir dos valores registados das diferentes variáveis disponibilizadas;
- Preparar séries de dados e executar a previsão de valores de produção de biogás durante a digestão anaeróbia;
- Comparar os resultados produzidos neste caso de estudo com os resultados da investigação realizada na ETAR do Seixal;
- Analisar os resultados e avaliar o potencial demonstrado pelas RNA na modelação do processo de digestão anaeróbia.

Concluída a investigação, definiu-se ainda um objetivo adicional que passou por gerar uma RNA que englobasse os dados disponíveis de cada um dos casos de estudo, de modo a se criar um modelo geral. Espera-se que o presente documento possa constituir-se como um

contributo valioso na aquisição de novos conhecimentos, no que diz respeito à modelação do processo de digestão anaeróbia, com recurso a RNA.

1.4 Organização da Dissertação

A dissertação está estruturada da seguinte forma:

- No primeiro capítulo é efetuada uma introdução à problemática em análise no trabalho e são dados a conhecer o enquadramento, os principais objetivos a cumprir na realização da dissertação e a forma como esta está estruturada;
- No segundo capítulo, é realizada uma revisão bibliográfica dos temas equacionados e com maior projeção na realização da dissertação, com especial incidência no processo de digestão anaeróbia e nas RNA;
- No terceiro capítulo, é dada informação acerca das entidades gestoras envolvidas nos estudo, são caracterizados as ETAR dos dois casos de estudo e é descrita a metodologia utilizada na execução da investigação;
- No quarto capítulo, são explicados os princípios básicos de funcionamento do programa de computador NeuralTools[®], usado na produção das RNA;
- No quinto capítulo são apresentados e discutidos os resultados da investigação efetuada na realização desta dissertação;
- No sexto capítulo são registadas as principais conclusões e sugerem-se recomendações para desenvolvimentos futuros no âmbito da atual dissertação.

2. Revisão Bibliográfica

2.1 Águas Residuais

Qualquer comunidade produz resíduos através das suas atividades, é inevitável que assim seja. Uma parte significativa destes resíduos acaba como água residual, a qual é gerada após a água de consumo ser utilizada em diversas aplicações (Henze e Comeau, 2008; Metcalf & Eddy, 2003). A água (H_2O) devido às suas propriedades de solvente e à capacidade de transporte que possui incorpora em si diversas impurezas, resultantes da sua utilização. (Von Sperling, 2007). A quantidade e qualidade da água residual produzida dependem de vários fatores, entre os quais o comportamento, nível e estilo de vida de cada população, bem como das condições técnicas e do quadro legal existentes de região para região (Henze e Comeau, 2008). A composição das águas residuais é assim resultado do tipo de uso a que a água foi sujeita, por parte de uma população (Von Sperling, 2007).

A água residual é caracterizada pelas suas propriedades físicas, químicas e biológicas, sendo que as principais podem ser observadas na Tabela 2.1 (Metcalf & Eddy, 2003).

Tabela 2.1 – Principais propriedades físicas, químicas e biológicas que compõem as águas residuais (adaptada de Metcalf & Eddy, 2003).

Águas Residuais		
Propriedades Físicas	Propriedades Químicas	Propriedades Biológicas
Condutividade;	<u>Inorgânicas:</u>	Coliformes;
Cor;	Alcalinidade;	Microorganismos (Bactérias,
Densidade;	Azoto Amoniacal;	Helminhas, Protozoários, Vírus);
Odor;	Azoto Orgânico;	Toxicidade.
Temperatura;	Cloretos;	
Teor em Sólidos Dissolvidos;	Fósforo Inorgânico	
Teor em Sólidos Suspensos;	Fósforo Orgânico;	
Teor em Sólidos Totais;	Fósforo Total;	
Transmitância;	Metais;	
Turvação.	Nitratos;	
	Nitritos;	
	pH;	
	Sulfatos.	
	<u>Orgânicas:</u>	
	Carência Bioquímica de	
	Oxigênio;	
	Carência Química de Oxigênio;	
	Carbono Orgânico Total.	

Quando as águas residuais não tratadas se acumulam e se tornam sépticas, ocorre a decomposição da matéria orgânica, gerando-se uma situação que pode ser bastante perigosa para a saúde humana e para os meios naturais. Nestas condições produzem-se gases fétidos e desenvolvem-se comunidades de microrganismos patogénicos comuns ao trato intestinal humano. Existe ainda a possibilidade de se gerarem episódios de eutrofização, devido a um estímulo do crescimento de plantas aquáticas e ao facto de poderem existir substâncias tóxicas ou compostos potencialmente mutagénicos e cancerígenos neste tipo de águas. Por estes motivos, existe a necessidade de recolher e drenar rapidamente as águas residuais, efetuar o seu tratamento e em seguida reutilizá-las ou devolve-las novamente ao meio-ambiente, tendo como objetivo a proteção da saúde pública e do ambiente. (Metcalf & Eddy, 2003).

2.2 Tratamento de Águas Residuais

A engenharia sanitária é o ramo da engenharia que tem a responsabilidade de desenvolver soluções técnicas e científicas, com capacidade de resolver os problemas associados ao tratamento das águas residuais (Metcalf & Eddy, 2003). Nos estudos de planeamento necessários à implementação de um sistema de tratamento de águas residuais, é essencial obter informação acerca dos seguintes aspetos:

- Objetivos de tratamento;
- Nível de tratamento necessário e eficiências de remoção a cumprir;
- Grau de sensibilidade da massa de água onde será descarregada a água após tratamento.

A legislação em vigor em cada país tem uma grande influência no padrão de qualidade a satisfazer, o qual tem um grande peso na definição dos aspetos mencionados. (Von Sperling, 2007). Os requerimentos legislativos a cumprir, em combinação com a qualidade do efluente e as características do terreno onde será implementada a ETAR, definem o tipo e o nível de tratamento a instalar (Qasim, 1999).

2.2.1 Legislação em Portugal

Em Portugal, no que diz respeito à legislação relativa ao tratamento de águas residuais, o primeiro regulamento adotado a nível nacional, foi a Diretiva 91/271/CEE do Conselho das Comunidades Europeias, relativa ao tratamento e descarga de águas residuais urbanas e descarga de águas residuais de determinados setores industriais. Esta diretiva foi executada, de modo a ditar as regras que os Estados Membro da Comissão Europeia deveriam respeitar, com o objetivo de proteger o ambiente dos efeitos negativos das descargas de águas residuais. Nesta diretiva são definidos os requisitos para as descargas das estações de tratamento de águas residuais (Tabela 2.2 e Tabela 2.3). Em ambos os casos serão aplicados os valores de concentração ou a percentagem mínima de redução, sendo apenas necessário cumprir um dos dois requisitos.

Tabela 2.2 – Requisitos para as descargas das estações de tratamento de águas residuais (Diretiva 91/271/CEE)

Parâmetros	Concentração (Requisito 1)	Percentagem de redução mínima (Requisito 2) ⁽¹⁾
Carência bioquímica de oxigénio (CBO ₅ a 20 °C) sem nitrificação ⁽²⁾	25 mg/L O ₂	70 – 90 40 nos casos previstos do artigo 4º
Carência química de oxigénio (CQO)	125 mg/L O ₂	75
Total de partículas sólidas em suspensão	35 mg/L ⁽³⁾	90 ⁽³⁾
	35 nos casos previstos no nº2 do artigo 4º (e.p superior a 10 000)	90 nos casos previstos no nº2 do artigo 4º (e.p superior a 10 000)
	60 nos casos previstos no nº2 do artigo 4º (e.p de 2000 a 10 000)	70 nos casos previstos no nº2 do artigo 4º (e.p de 2000 a 10 000)

(1) Redução em relação à carga de afluente.

(2) O parâmetro pode ser substituído por outro: carbono orgânico total (COT) ou carência total de oxigénio (CTO), se for possível estabelecer uma relação entre a CBO₅ e o parâmetro de substituição.

(3) Este requisito é facultativo.

Tabela 2.3 – Requisitos para as descargas de águas residuais urbanas em zonas sensíveis sujeitas a eutrofização (Diretiva 91/271/CEE)

Parâmetros	Concentração (Requisito 1)	Percentagem de redução mínima (Requisito 2) ⁽¹⁾
Fósforo Total	2 mg/L P (10 000 – 100 000 e.p.) 1 mg/L P (mais de 100 000 e.p.)	80
Azoto total ⁽²⁾	15 mg/L N (10 000 – 100 000 e.p.) 10 mg/L N (mais de 100 000 e.p.) ⁽³⁾	70 - 80

(1) Redução em relação à carga de afluente.

(2) Por azoto total entende-se a soma do total de azoto-Kjeldahl (N orgânico+NH₃), azoto de nitratos (NO₃) e azoto de nitritos (NO₂). Este requisito é facultativo.

(3) Alternativamente, a média diária não poderá exceder 20 mg/L N. Este requisito refere-se a uma temperatura de água igual ou superior a 12 °C, durante o funcionamento do reator biológico da instalação de tratamento de águas residuais. Em substituição do critério da temperatura poderá utilizar-se um critério de limitação do tempo de funcionamento que tenha em conta as condições climáticas locais. Esta alternativa aplica-se no caso em que seja possível demonstrar que se cumpre o disposto na alínea D), nº 1, do anexo I, da Diretiva 91/271/CEE do Conselho, de 21 de Maio.

O Decreto-Lei nº 152/97 de 19 de Junho transpõe para o direito interno a Diretiva 91/271/CEE, do Conselho, de 21 de Maio de 1991, no que diz respeito ao tratamento de águas residuais urbanas. Os requisitos instituídos pela Diretiva 91/271/CEE em relação às descargas das estações de tratamento de águas residuais, inclusive nas zonas sensíveis não foram alterados na transposição para o Decreto-Lei nº 152/97, estando ainda em vigor no presente. No dia 9 de Novembro de 1998 é aprovado o Decreto-Lei nº 348/98 e no dia 22 de Junho de 2004 é

aprovado o Decreto-Lei nº 149/2004, os quais fazem pequenas alterações ao Decreto-Lei nº 152/97, principalmente no que diz respeito à identificação das zonas sensíveis para descargas.

Além da legislação implementada como instrumento direto de regulação e gestão das águas residuais urbanas, há que realçar outros documentos que têm tido grande influência nesta área. No ano de 2000 é aprovado o Plano estratégico de Abastecimento de Água e de Saneamento de Águas Residuais (PEAASAR). Este plano foi elaborado para vigorar entre os anos de 2000 a 2006, tendo sido reconhecida a sua importância nos progressos verificados durante este período no âmbito do abastecimento de água e de saneamento de águas residuais. Apesar de os progressos efetuados, subsistiam problemas por resolver, o que aliado ao novo ciclo de fundos comunitários (2007-2013), consubstanciado pelo Quadro de Referência Estratégico Nacional (QREN), justificaria uma nova estratégia (PEAASAR II, 2007).

Em 2007 é dado a conhecer o PEAASAR II. Um dos principais pressupostos da base de conceção deste plano foi a articulação com outros documentos de relevo nacional no domínio dos recursos hídricos, entre os quais se destacam (PEAASAR II, 2007):

- A Lei da Água (Decreto-Lei nº 58/2005, de 29 de Dezembro) que transpôs para o direito nacional a Diretiva Quadro de Água (DQA), a qual é desde a data da sua elaboração (2000) o principal instrumento da Política da União Europeia, relativa à água;
- A Lei da Titularidade dos Recursos Hídricos (Decreto-Lei nº 54/2005, de 15 de Novembro);
- O Plano Nacional da Água (Decreto-Lei nº 112/2002, de 17 de Abril).

O PEAASAR II definiu novos objetivos e medidas de otimização, a serem implementadas durante os anos de 2007 a 2013, medidas essas que visaram reforçar as infraestruturas do setor, promovendo a sustentabilidade e a continuidade e qualidade dos serviços prestados. (PEENSAAR 2020, 2015) O balanço efetuado no final deste plano, em relação ao saneamento de águas residuais apresentou aspetos positivos e negativos. As medidas adotadas permitiram um aumento bastante significativo na qualidade das águas balneares e um aumento notório da qualidade das restantes massas de água. Em contrapartida o objetivo operacional principal não foi cumprido, dado que no ano de 2013, a percentagem de população servida com sistemas públicos de abastecimento era inferior a 90% (PEENSAAR 2020, 2015). Durante o período em que o PEAASAR II esteve em vigor foi promulgado o Decreto-Lei nº 198/2008, de 8 De Outubro, o qual alterou o Decreto-Lei nº 152/97, de 19 de Junho, na redação que lhe foi dada pelo Decreto-Lei nº 348/98, de 9 de Novembro e o Decreto-Lei nº 149/2004, de 22 de Junho.

No dia 30 de Abril de 2015, foi aprovado através do Despacho nº 4385/2015 do Gabinete do Secretário de Estado do Ambiente, o “PENSAAR 2020 - Uma nova estratégia para o setor de abastecimento de águas e saneamento de águas residuais”. Este documento é resultado do balanço do PEAASAR II, do diagnóstico da situação de referência e da visão, objetivos, metas e cenários que definem o quadro estratégico para o período entre os anos de 2014 e 2020. O

novo plano tem como visão “Um setor ao serviço da população e da economia do País que presta serviços de qualidade e sustentáveis em termos ambientais, económico-financeiros e sociais”. Para sustentar a nova visão, são definidos no PENSAAR 2020 os seguintes objetivos estratégicos (PENSAAR 2020, 2015):

- Proteger o ambiente e melhorar da qualidade das massas de água;
- Melhorar a qualidade dos serviços prestados;
- Otimizar e efetuar uma gestão eficiente dos recursos;
- Garantir a sustentabilidade económico-financeira e social;
- Assegurar as condições básicas para que os problemas transversais ao setor possam ser resolvidos.

A partir destas linhas de ação, espera-se que nos próximos anos haja uma focalização na gestão dos ativos, no seu funcionamento e na qualidade de serviços prestada. Este enfoque irá justificar o esforço realizado pelo País nos últimos anos, no reforço da infraestruturação afeta ao abastecimento de água e saneamento de águas residuais (PENSAAR 2020, 2015).

Em relação à legislação, entrou recentemente em vigor o Decreto-Lei nº 133/2015, de 13 de Julho, que alterou o Decreto-Lei nº 152/97, de 19 de Junho. As respetivas modificações verificadas dizem respeito à classificação e delimitação de zonas sensíveis no território português. Constatase que ao longo dos anos as mudanças registadas nos Decretos-Lei, referem-se sempre à classificação das zonas sensíveis, o que vem atender à necessidade de realizar revisões periódicas destas zonas, algo que foi imposto pela Diretiva nº 91/271/CEE, do Conselho, de 21 de Maio de 1991.

No âmbito do saneamento de águas residuais há que realçar, que todos os documentos mencionados têm sido de extrema importância, no apoio ao cumprimento da legislação e regulamentos, afetos ao setor.

2.2.2 Estações de Tratamento de Águas Residuais

Dentro das estruturas onde se efetua o tratamento de águas residuais, existem vários processos de tratamento, instalados consoante o nível de tratamento necessário. Na escolha dos processos e equipamentos, existem muitos fatores a considerar como os constituintes e limitações do efluente a tratar, proximidade da área urbana, perfil do terreno, requisitos energéticos e o orçamento disponível para a construção da ETAR (Qasim, 1999). No projeto de uma estação de tratamento de águas residuais é necessário garantir, que o efluente tratado se encontra dentro dos parâmetros definidos pelo quadro legal e que este não afeta negativamente a massa de água, onde é descarregado. Para o cumprimento dos objetivos propostos numa ETAR encontramos habitualmente, duas fases de tratamento diferentes, ou

seja, uma linha de tratamento da fase líquida e uma linha de tratamento da fase sólida (Vesilind, 2003).

Linha de Tratamento da fase líquida

Por norma, a linha de tratamento da fase líquida detém quatro níveis de tratamento distintos (Tabela 2.4).

Tabela 2.4 – Níveis de tratamento de águas residuais (adaptada de Metcalf & Eddy, 2003; Von Sperling, 2007)

Nível	Capacidade de Tratamento
Preliminar	Remoção de sólidos, areias, óleos e gorduras,
Primário	Remoção de sólidos suspensos e de matéria orgânica.
Secundário	Remoção de sólidos suspensos e de matéria orgânica biodegradável (solúvel e em suspensão).
Terciário ou Avançado	Remoção da restante parte de sólidos suspensos, de nutrientes, de metais, de compostos não biodegradáveis, de compostos inorgânicos não dissolvidos e de organismos patogênicos.

Quando a água bruta chega à obra de entrada de uma ETAR, primeiro passa por um certo número de operações exclusivamente físicas e mecânicas que compõem o tratamento preliminar. De acordo com a dimensão do caudal recebido e a qualidade do mesmo, pode existir uma ou mais das seguintes operações (Degremont, 2002):

- Gradagem grossa – Esta operação remove os sólidos grosseiros e outros detritos de dimensões elevadas existentes na água bruta, de modo a evitar obstruções nos equipamentos e estruturas de tratamento localizadas a jusante;
- Gradagem fina - Remove sólidos finos e pequenas partículas para que estes não afetem as eficiências de tratamento de processos seguintes;
- Desarenação – Esta etapa tem como objetivo remover a areia presente no efluente, de forma a prevenir a deposição e acumulação desta noutros equipamentos;
- Flotação – Aplica-se esta etapa quando é necessário remover óleos e gorduras, com o intuito de evitar entupimentos e outros problemas operacionais em processos de tratamento posteriores.

O efluente após atravessar as operações correspondentes ao tratamento preliminar é conduzido até ao tratamento primário. Quase todas as plantas de tratamento incluem nesta fase, um processo de decantação como etapa de tratamento. O propósito da decantação é a remoção de uma parte substancial de matéria orgânica e de sólidos suspensos existentes no

efluente, que de outra maneira seriam descarregados diretamente num curso de água. A legislação portuguesa de acordo com o Decreto-Lei nº 152/97, de 19 de Junho, obriga a remover pelo menos 20% de CBO_5 e 50% das partículas sólidas em suspensão – independentemente do processo físico e/ou químico que envolva a decantação, ou qualquer outro processo de igual capacidade. O rigor dos requisitos legislativos a cumprir e o tamanho da instalação definem geralmente o número de decantadores a instalar e as suas dimensões (Metcalf & Eddy, 2003).

Finalizado o tratamento primário, segue-se o tratamento secundário. Os principais objetivos nesta etapa são a remoção da matéria orgânica solúvel e coloidal que não é possível eliminar através do tratamento primário, a remoção de nutrientes como o fósforo (P) e o azoto (N) e a remoção dos sólidos que ainda se encontram no efluente. Face aos desafios existentes, o mais comum nesta etapa é o recurso ao tratamento por processos biológicos, onde os microrganismos desempenham o papel principal, mas também é possível recorrer a processos químicos ou a sistemas naturais (Qasim, 1999). Existem vários diagramas possíveis de aplicar no tratamento secundário, mas os mais típicos são as lamas ativadas, as lagoas de estabilização, os leitos percoladores, os biofiltros e os discos biológicos (Metcalf & Eddy, 2003). O tratamento secundário tem um grande peso no cumprimento da legislação, nomeadamente dos requisitos presentes no quadro nº 1 e nº 2 do Decreto-Lei nº 152/97, de 19 de Junho (Tabela 2.2 e Tabela 2.3).

Por último na linha de tratamento da fase líquida ocorre o tratamento terciário ou o tratamento avançado. Um dos objetivos deste nível de tratamento é a remoção complementar de sólidos suspensos, de nutrientes, de metais, de compostos não biodegradáveis e de compostos inorgânicos não dissolvidos, caso estas substâncias ainda estejam presentes em proporções que justifiquem a sua eliminação. Em caso afirmativo, as tecnologias mais comuns que se costumam utilizar na sua supressão são os processos de filtração por membrana (Metcalf & Eddy, 2003). Outro dos objetivos principais do tratamento terciário é erradicar os organismos patogénicos presentes no efluente (Qasim, 1999). Os métodos de desinfecção que podem ser usados estão divididos em quatro categorias: físicos, químicos, mecânicos e por radiação (Metcalf & Eddy, 2003). Por norma, o cloro e os seus compostos são o método de desinfecção mais usado em todo o mundo, apesar de a radiação por ultravioleta também ser bastante usada (Qasim, 1999).

Linha de Tratamento da fase sólida

Quando as águas residuais são tratadas de forma a retornarem ao meio-ambiente, geram-se sólidos ricos em nutrientes, vulgarmente apelidados de lamas, compostas por 0,25% a 7% de fração sólida (Turovskiy e Mathai, 2006). Dos constituintes extraídos ao longo do tratamento da fase líquida, as lamas são de longe a substância que gera um maior volume acumulado de

resíduos, constituindo um dos problemas mais complexos de resolver no tratamento de águas residuais (Metcalf & Eddy, 2003). A quantidade de lamas normalmente tratada representa apenas 1% da quantidade de águas residuais tratadas, mas enquanto o processo de tratamento da fase líquida demora horas até se completar, o tratamento de lamas só é finalizado ao fim de vários dias ou mesmo semanas (Turovskiy e Mathai, 2006). Os problemas relacionados com o tratamento de lamas são difíceis de resolver, devido a estas serem na sua maioria compostas por substâncias responsáveis pelo carácter perigoso das águas residuais não tratadas. (Metcalf & Eddy, 2003). As características e quantidade das lamas geradas são muito variáveis, dependendo de forma direta da qualidade do efluente e do tipo de tratamento a que este foi sujeito na linha líquida (Vesilind, 2003).

O principal propósito do tratamento da fase sólida é a redução de volume e a estabilização biológica das lamas. A redução de volume pode ser conseguida através de processos de espessamento, de desidratação e de incineração. As lamas podem ser estabilizadas a partir de processos como a digestão anaeróbia, a digestão aeróbia, a estabilização química e a incineração (Degrèmont, 2002).

Usualmente o primeiro processo da linha de tratamento da fase sólida é o espessamento de lamas, que ocorre habitualmente por espessamento gravítico, flotação ou centrifugação, dependendo da qualidade de lamas afluentes. A etapa de espessamento permite aumentar a concentração das lamas através da remoção de parte da fração líquida (Metcalf & Eddy, 2003).

Se as lamas forem consideradas inorgânicas (compostas por uma percentagem de matéria orgânica <30%), normalmente não necessitam de serem estabilizadas. Caso contrário, antes da etapa de desidratação, é comum existir um processo de estabilização biológica de lamas (Degrèmont, 2002).

O principal objetivo da estabilização de lamas é a redução de organismos patogénicos, eliminação de odores e controlo do potencial de putrefação da matéria orgânica. Atualmente, a capacidade de produzir metano (biogás) também é muito apreciada, tornando-se uma vantagem adicional (Metcalf & Eddy, 2003). Nos últimos anos a digestão anaeróbia tem sido o processo de estabilização mais usado nas estações de tratamento de média e grande dimensão, por possuir uma boa eficiência energética e necessitar de uma reduzida quantidade de químicos (Qasim, 1999).

No fim da linha habitualmente, está instalada uma operação física de desidratação, de modo a reduzir mais uma vez o volume das lamas. A redução de volume nesta fase é particularmente importante pela redução dos custos associados ao armazenamento e transporte das lamas a destino final. A desidratação também aumenta o poder calorífico das lamas - o que é vantajoso caso as lamas sigam para incineração - e reduz os problemas de acondicionamento das

mesmas, se estas forem enviadas para compostagem (Metcalf & Eddy, 2003). A seleção da tecnologia a empregar na desidratação depende das características das lamas e do espaço disponível na instalação. Quando existe área disponível e a quantidade de lamas a desidratar não é elevada, sistemas naturais como os leitos de secagem são opções bastante válidas. Perante locais com menos espaço disponível, os sistemas mecânicos são a melhor opção, ou seja as centrífugas, os filtros prensa, os filtros banda e os filtros de vácuo (Qasim, 1999).

Segundo o PENSAAR 2020, o destino final das lamas em Portugal é quase sempre a valorização agrícola, no entanto pretende-se analisar e criar nos próximos anos novas soluções integradas para a gestão de lamas, promovendo mais-valias do ponto de vista económico e ambiental. Um dos objetivos operacionais “imperiosos” do PENSAAR 2020 é o aumento da ecoeficiência nas ETAR, através da valorização das lamas com valor para outros usos e da gestão e produção de energia. Considera-se que a energia elétrica consumida no saneamento de águas residuais pode ser minimizada, através do aproveitamento do potencial energético das lamas, no processo de digestão anaeróbia para produção de biogás. Perante a melhoria pretendida no desempenho energético das ETAR, afirma-se também que no atual contexto económico-ambiental, se deve mudar o paradigma de visão das lamas como um resíduo para uma visão de recurso valorizável e estudar formas de otimizar o rendimento dos equipamentos e processos.

2.3 Processo de Digestão Anaeróbia

As tecnologias de tratamento baseadas em processos anaeróbios têm sido usadas um pouco por todo o mundo, principalmente no tratamento de águas residuais e de lamas municipais e no tratamento de vários tipos de águas residuais industriais. Os sistemas anaeróbios são atrativos devido ao facto de requererem menos energia que as tecnologias aeróbias. Em adição, são ainda capazes de produzir energia sobre a forma de biogás, a qual pode ser convertida em energia mecânica ou elétrica e ser usada na própria estação de tratamento (Malina e Pohland, 1992).

Dos processos anaeróbios existentes, a digestão anaeróbia é um dos mais antigos e amplamente usados em engenharia sanitária na degradação e estabilização de matéria orgânica, sobretudo associado ao tratamento de águas residuais urbanas e industriais (Metcalf & Eddy, 2003). O principal propósito deste processo é a conversão das lamas até à mineralização e facilitar a desidratação numa etapa posterior. Durante o processo que ocorre na ausência de oxigênio, uma parte da matéria orgânica existente é microbiologicamente convertida a biogás (Malina e Pohland, 1992).

Os primeiros desenvolvimentos da tecnologia de digestão anaeróbia remontam à segunda metade do século XIX, mas na verdade este é um processo natural que se sucede quando o

oxigênio não está disponível (Kythreotou *et al*, 2014). Apesar de desde o século XVII haver relatos que a decomposição de matéria orgânica gerava gases inflamáveis, só em 1881 é dado o primeiro passo no desenvolvimento da tecnologia de digestão anaeróbia, através da aplicação de uma fossa séptica ainda em estado embrionário, no tratamento de águas residuais domésticas. Durante a maior parte do século XX, os desenvolvimentos deste processo mantiveram-se exclusivamente ligados à estabilização de matéria orgânica putrefaciente de águas residuais domésticas. A aplicação de sistemas de digestão anaeróbia de forma mais abrangente foi estimulada a partir da década de 70, devido à subida do preço dos combustíveis fósseis e à publicação de normas de poluição mais rigorosas (Abbasi *et al*, 2012). Nas últimas décadas, este processo tem sido amplamente estudado também para aplicação no tratamento de resíduos sólidos, no tratamento de águas residuais industriais e na estabilização dos resíduos da biomassa agrícola ou florestal (Lindmark *et al*, 2014). No presente, a digestão anaeróbia tem sido extensivamente usada não só por causa do tratamento de lamas, mas também devido à capacidade que possui para produção de biogás. Previsões apontam para que o número de reatores venha a aumentar novamente, dada a aposta nas energias renováveis, nomeadamente no biogás e dado o relevo que as alterações climáticas têm tido (Donoso-Bravo *et al*, 2011).

2.3.1 Vantagens e Desvantagens da Digestão Anaeróbia

Entre as vantagens que a digestão anaeróbia oferece em relação a outros processos de estabilização de lamas, estão as seguintes (Malina e Pohland, 1992; Turovskiy e Mathai, 2006):

- A produção de biogás é uma fonte de energia viável. Se a estação de tratamento for de dimensão considerável, em muitos casos, a energia gerada supera a energia despendida na manutenção da temperatura no digestor anaeróbio e da manutenção dos dispositivos de mistura. O excesso de energia pode assim ser usado noutros requisitos energéticos das instalações adjacentes;
- Redução da massa total de lamas, através da conversão da matéria orgânica essencialmente em biogás e água. Em situações normais, 30% a 65% da matéria sólida bruta afluente é destruída, o que significa uma redução nos custos do restante tratamento e eliminação de lamas;
- As lamas digeridas habitualmente estão livres de odores desagradáveis;
- As lamas biológicas digeridas contêm uma porção de matéria orgânica e nutrientes como o azoto e o fósforo que aumentam a fertilidade dos solos, se estas forem usadas na agricultura;
- Uma elevada taxa de eliminação de organismos patogénicos pode ser alcançada, especialmente no funcionamento do digestor anaeróbio em gama termófila.

Os processos de tratamento não são perfeitos tal como a tecnologia de digestão anaeróbia, a qual apresenta como desvantagens os seguintes aspetos (Malina e Pohland, 1992; Turovskiy e Mathai, 2006):

- Os custos de capital são elevados, porque o funcionamento de um digestor anaeróbio requer sistemas de alimentação, aquecimento e mistura de lamas que apresentam altos custos de manutenção;
- São necessários tempos de retenção hidráulica longos para estabilização de lamas. O processo de digestão é relativamente lento, pelo que limita a velocidade de reação com que o sistema se consegue adaptar a alterações nas condições de funcionamento do mesmo;
- Os microrganismos envolvidos na digestão anaeróbia são sensíveis a pequenas alterações nas condições de funcionamento do reator. O processo é assim vulnerável a problemas de operação, as quais para ser combatidas requerem um controlo apertado de todos os requisitos de funcionamento do digestor;
- O processo produz sobrenadantes de baixa qualidade com grande carência de oxigénio e altas concentrações de sólidos suspensos, azoto e fósforo. Estes sobrenadantes necessitam muitas vezes de tratamento adicional, antes de o caudal ser reenviado para os equipamentos que removem nutrientes das águas residuais.

2.3.2 Teoria do Processo de Digestão Anaeróbia

A Digestão anaeróbia é um processo complexo que requer condições anaeróbias estritas e depende de uma atividade coordenada entre as culturas de microrganismos, para que a conversão da matéria orgânica na sua maioria em biogás ocorra eficazmente. O processo envolve diferentes estágios de reações químicas e biológicas, sendo as principais fases deste processo a hidrólise, a acidogénese, a acetogénese e a metanogénese (Figura 2.1) (Appels *et al*, 2008; Yu *et al.*, 2013).

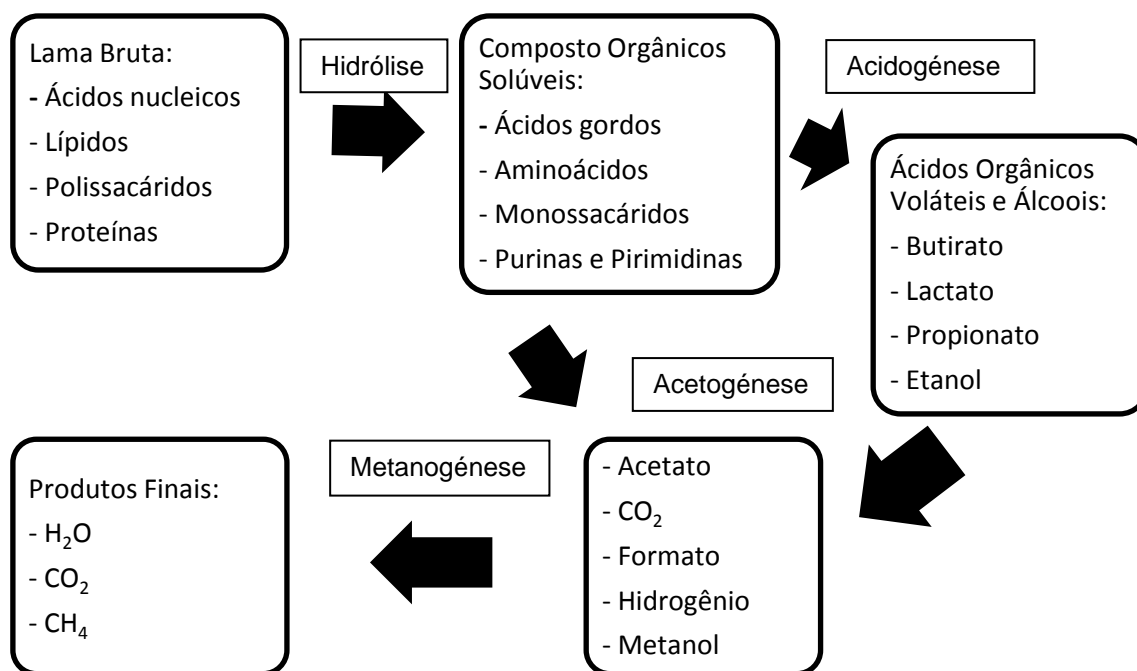


Figura 2.1 – Diagrama simplificado das reações presente no processo de digestão anaeróbia (adaptado de Lapa, 2014; Turovskiy e Mathai, 2006)

Hidrólise

Durante hidrólise ocorre a degradação de sólidos voláteis de peso molecular elevado, como os ácidos nucleicos, lípidos, polissacáridos e proteínas. Estes polímeros complexos são, pela ação de um consórcio de bactérias hidrolíticas, transformados em monómeros de forma solúvel como os ácidos gordos, aminoácidos, monossacáridos e as purinas (Turovskiy e Mathai, 2006). Nesta fase os compostos são catalisados por enzimas extracelulares, por exemplo as proteases são a enzima que serve de catalisador para a conversão das proteínas, enquanto as lipases atuam da mesma maneira em relação aos lípidos. As enzimas produzidas pelas bactérias são depois excretadas pelas mesmas para a fase líquida (Lin Chen *et al*, 2014).

A extensão da estabilização anaeróbia é avaliada segundo a quantidade de sólidos voláteis destruídos. Uma percentagem de eliminação inferior a 40% indica que existem concentrações elevadas de matéria difícil de degradar. Percentagens de eliminação superiores a 65% indicam que as lamas contêm substâncias fáceis de degradar como as que são degradadas pelas bactérias hidrolíticas. Em situações em que a matéria de difícil degradação é demasiado abundante, a hidrólise pode ser considerada como a fase limitante do processo de digestão anaeróbia (Vesilind, 2003).

Acidogénese

No decorrer da acidogénese dá-se a fermentação dos compostos orgânicos solúveis e respetiva conversão em ácidos orgânicos voláteis (butirato, lactato, entre outros), álcoois (etanol, metanol, entre outros), acetato, CO₂ e hidrogénio (H₂) (Lin Chen *et al*, 2014). As bactérias acidogénicas responsáveis pelo processo são normalmente facultativas (apresentam atividade em ambientes com e sem oxigénio), apesar de existirem também bactérias estritamente anaeróbias. Estas bactérias costumam apresentar um crescimento rápido e são relativamente tolerantes a mudanças nas condições de operação do digestor, como alterações de pH e de temperatura. Nesta etapa, não há praticamente alterações na quantidade de matéria orgânica presente no sistema, contudo existe uma redistribuição dos vários tipos de compostos orgânicos simples (Malina e Pohland, 1992).

Acetogénese

A acetogénese é um processo realizado no interior das células, onde entram em ação as bactérias homoacetogénicas e as bactérias sintróficas (produtoras obrigatórias de H₂). As homoacetogénicas, a partir dos ácidos orgânicos voláteis (AOV), dos ácidos gordos, dos aminoácidos, dos monossacáridos e das purinas e pirimidinas produzem H₂, CO₂, acetato, formato e metanol (Lapa, 2014).

As reações que se geram na produção de H₂ são termodinamicamente desfavoráveis, se não existir uma quantidade suficiente de microrganismos com capacidade de usar o hidrogénio (Degrèmont, 2002). Perante concentrações de H₂ em excesso, e a sua não remoção, este estágio pode ser bloqueado (Lapa, 2014). No entanto, a partir do momento em que a concentração de H₂ no meio é igual ou inferior a 10⁻⁴ atm, as reações tornam-se termodinamicamente favoráveis e a variação de energia é suficiente para permitir a síntese de trifosfato de adenosina (ATP) e o crescimento das bactérias (Degrèmont, 2002; Lapa, 2014).

Metanogénese

A metanogénese é considerada por vários autores de forma unânime como a etapa chave da digestão anaeróbia. Nesta fase ocorre a degradação dos substratos com baixo número de átomos de carbono, como o acetato, o formato, o metanol, o CO₂ e o H₂ (Lapa, 2014). Existem dois tipos de reações maioritárias na metanogénese, a partir das quais se formam o CH₄ (produto principal), o CO₂ e H₂O. A reação entre o H₂ e o CO₂ gera H₂O e CH₄ (equação 1.1), enquanto a quebra das moléculas de acetato dá origem ao CO₂ e ao CH₄ (equação 1.2) (Degrèmont, 2002; Lapa, 2014).



As bactérias metanogénicas são anaeróbias obrigatórias e têm também um papel fundamental na acetogénese, através da remoção do excesso de H_2 presente no meio, que caso não acontecesse, afetaria a atividade das bactérias sintróficas, podendo, em última instância, bloquear a acetogénese (Lapa, 2014). Para um bom funcionamento do digestor anaeróbio, é fundamental que a acidogénese e a acetogénese estejam em equilíbrio dinâmico com a metanogénese. Após a conversão dos compostos orgânicos em AOV e H_2 é essencial que ocorra ao mesmo ritmo a formação de CH_4 e de CO_2 . Como resultado é normal que a concentração dos AOV e do H_2 seja baixa num reator a funcionar bem. No entanto, o bom funcionamento da metanogénese é suscetível de ser afetado, visto que as bactérias metanogénicas são de crescimento lento e a sua atividade pode ser ameaçada por alterações no pH e na temperatura dentro do digestor. Em contraste, a acidogénese e acetogénese são processos mais rápidos e as suas bactérias aguentam melhor as mudanças das condições de funcionamento. Quando se verificam várias oscilações no ambiente do reator, as bactérias metanogénicas podem não conseguir gerar CH_4 à velocidade de formação dos AOV e do H_2 , o que cria um desequilíbrio. No caso de o equilíbrio dinâmico ser afetado, a metanogénese pode ser inibida e originar o colapso de todo o sistema se nenhuma medida corretiva for tomada (Turovskiy e Mathai, 2006).

2.3.3 Tipos de Digestão Anaeróbia

Na prática, existem essencialmente dois tipos principais de digestão anaeróbia que têm sido desenvolvidas ao longo dos anos, a digestão em baixa carga e a digestão em alta carga. Existem outros tipos de digestão como a digestão anaeróbia em duas fases e a digestão anaeróbia em dois estágios, contudo, devido a algumas diferenças teóricas encontradas na literatura consultada e ao facto desta temática não ser objeto principal do estudo, estas não serão abordadas (Metcalf & Eddy, 2003; Qasim, 1999; Turovskiy e Mathai, 2006; Vesilind, 2003). Em adição, o digestor anaeróbio pode ser operado em três regimes de temperatura diferentes: na gama psicrófila, na gama mesófila e na gama termófila (Turovskiy e Mathai, 2006; Vesilind, 2003).

Digestão em baixa carga

A digestão em baixa carga é o modo mais antigo e simples de o digestor operar, o qual é alimentado de forma intermitente neste modelo de digestão. Durante este processo, normalmente não é promovido qualquer tipo de mistura nem nenhuma fonte de aquecimento externa, no entanto, o biogás gerado e que sobe até à superfície acaba por produzir alguma agitação e mistura dentro do digestor. No interior do digestor ocorre a estratificação de diferentes camadas (Figura 2.2). O sobrenadante produzido é comumente enviado para a cabeça da estação, enquanto as lamas estabilizadas se acumulam no fundo do digestor, de modo a serem periodicamente purgadas. O tempo de retenção de lamas na digestão em baixa

carga é normalmente bastante elevado, mais concretamente entre os 30 dias e os 60 dias (Qasim, 1999; Turovskiy e Mathai, 2006; Vesilind, 2003). Na maior parte dos casos, este modo de digestão é usado somente em ETAR pequenas que no máximo recebem um caudal na ordem dos 4000 m³/dia (Turovskiy e Mathai, 2006; Vesilind, 2003).

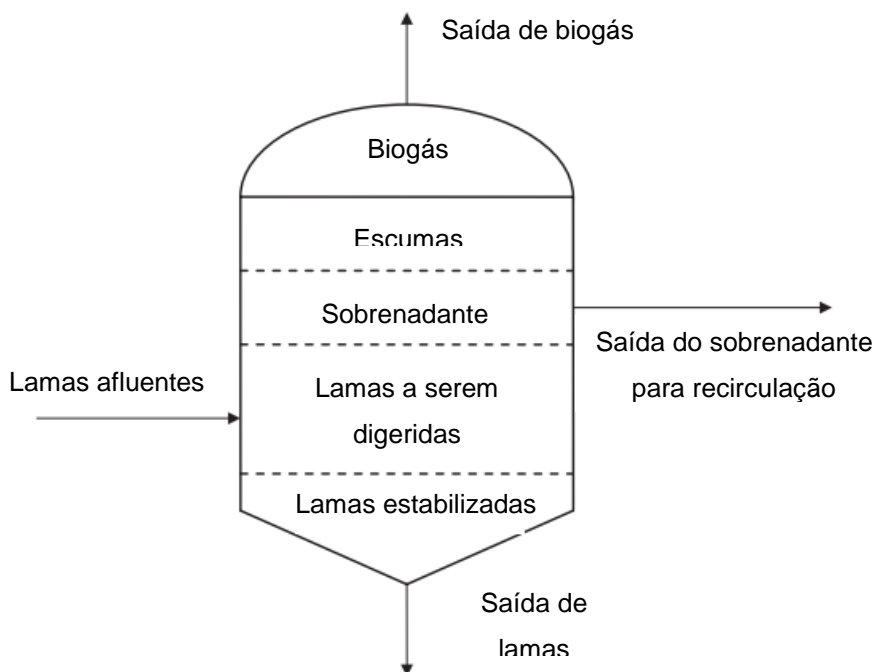


Figura 2.2 – Digestão anaeróbia em baixa carga (adaptada de Appels et al, 2008)

Digestão em alta carga

Na década de 50 foram realizados diversos melhoramentos na digestão de baixa carga, o que resultou no desenvolvimento de um novo modelo denominado como digestão em alta carga. O sistema de alta carga é caracterizado por ter equipamentos suplementares de aquecimento e mistura, por ser alimentado uniformemente e por conter uma etapa de espessamento antes da digestão, criando assim um ambiente o mais uniforme possível. Tendo em consideração estas propriedades, as principais vantagens do sistema de alta carga face ao de baixa carga são as seguintes (Turovskiy e Mathai, 2006):

- A estabilidade e a eficiência do processo aumentam;
- O volume efetivo do reator aumenta, o que permite realizar a digestão em digestores menores;
- A estratificação dentro do digestor é mais reduzida;
- As substâncias inibitórias são diluídas mais facilmente;
- A matéria inorgânica que tem tendência para assentar é mantida em suspensão com mais facilidade;
- O acumular de sobrenadante é reduzido;
- As reações de produção de biogás são facilitadas.

O tempo de retenção habitual das lamas em digestão neste sistema é de 10 a 20 dias (Figura 2.3) (Qasim, 1999).

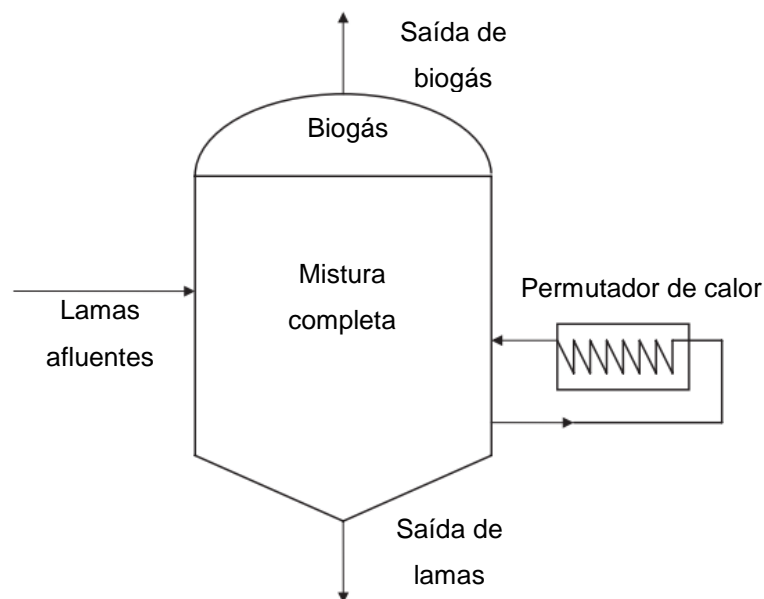


Figura 2.3 – Digestão anaeróbia em alta carga (adaptada de Appels et al, 2008)

Digestão nas diferentes gamas de temperatura

A etapa de digestão anaeróbia de lamas ocorre sobretudo em dois regimes de temperatura diferentes, o mesófilo entre os 30°C e os 38°C e a termófilo entre os 50°C e os 57°C, sendo que os digestores de alta carga operam normalmente às temperaturas mesófilas (Metcalf & Eddy, 2003). A digestão em regime de temperatura psicrófila também existe, no entanto costuma ser preterido em relação aos outros dois regimes, dado o funcionamento do digestor ser influenciado pelas condições climáticas e a quantidade de sólidos voláteis destruídos ser menor. Em geral, as vantagens do reator trabalhar na gama de temperatura termófila em comparação em relação à mesófila são as seguintes (Turovskiy e Mathai, 2006):

- As reações sucedem de forma mais veloz, o que aumenta a destruição dos sólidos voláteis;
- A eliminação de organismos patogénicos aumenta.

As desvantagens de operar em gama termófila face à gama mesófila são as seguintes (Turovskiy e Mathai, 2006):

- Requisitos energéticos mais elevados;
- A qualidade do sobrenadante é menor, porque possui amplas quantidades de materiais dissolvidos;
- Potencial de odor maior;
- Existe menos estabilidade no processo, devido ao facto de as bactérias termófilas serem mais sensíveis a flutuações de temperatura.

Os digestores termófilos têm sido usados apenas em determinadas aplicações no tratamento de lamas municipais, enquanto os digestores mesófilos são utilizados num maior número de casos. (Metcalf & Eddy, 2003).

2.3.4 Dimensionamento do Sistema de Digestão Anaeróbia

Alguns dos fatores mais importantes que influenciam o dimensionamento do sistema e que definem as condições de operação para as diferentes fases da digestão anaeróbia são o dimensionamento do digestor, o sistema de aquecimento e as condições de mistura. Todas estas propriedades do sistema estão interligadas entre si (Qasim, 1999).

Dimensionamento do Digestor

Os reatores são regularmente cilíndricos, podendo assumir formas paralelepípedicas ou em forma de ovo e as suas dimensões podem ser bastante variáveis. Em relação aos cilíndricos, o fundo é usualmente cónico com um declive mínimo de 1:6 (vertical para horizontal), com a cavidade de saídas das lamas localizada ao centro (Metcalf & Eddy, 2003). A determinação do volume do digestor anaeróbio é o primeiro aspeto a considerar na definição do sistema. O primeiro fator em ter conta no dimensionamento são as características quantitativas e qualitativas das lamas afluentes, pois estas influenciam as condições de operação. Vários métodos de cálculo podem ser usados no dimensionamento do volume do digestor, entre os quais estão a carga de sólidos, o equivalente populacional, o tempo de retenção de sólidos e a redução de volume (Turovskiy e Mathai, 2006). Na Tabela 2.5 é possível observar um resumo dos critérios de dimensionamento da etapa de digestão para processos em baixa carga e alta carga.

Tabela 2.5 – Critérios de dimensionamento para digestores anaeróbios (adaptado de Qasim, 1999)

Parâmetros	Baixa carga	Alta carga
Carga de sólidos voláteis (Kg/m ³ .dia)	0,64 – 1,60	2,40 – 6,41
Lamas primárias (m ³ /capita)	0,03 – 0,04	0,02 – 0,03
Lamas primárias + lamas ativadas em excesso (m ³ /capita)	0,06 – 0,08	0,02 – 0,04
Lamas primárias + lamas de leitos percoladores (m ³ /capita)	0,06 – 0,14	0,02 – 0,04
Tempo de retenção de sólidos (dias)	30 - 60	10 - 20

Um dos métodos mais comuns e diretos na definição do volume do digestor é a carga de sólidos voláteis. Geralmente, este procedimento é fundamentado na alimentação de lamas, sendo tipicamente considerado o valor máximo de lamas afluente do mês com as duas semanas de maior produção de sólidos (Metcalf & Eddy, 2003). O método mais antigo e simples é o equivalente populacional, o qual é baseado no número de população servida pela

ETAR, uma vez que a cada habitante é atribuído um certo volume (m^3 per capita). Este modo de cálculo deve ser utilizado apenas como uma predeterminação do volume, devido a estimar vários parâmetro que podem variar de comunidade para comunidade (Turovskiy e Mathai, 2006). O volume do digestor também pode ser calculado com base no tempo de retenção de sólidos, desde que o processo seja uma função do tempo que os microrganismos necessitam para digerir a matéria orgânica e se reproduzirem. Se a digestão anaeróbia não possuir recirculação de lamas ou eliminação de sobrenadante, o tempo de retenção de sólidos é igual ao tempo de retenção hidráulico (Appels *et al*, 2008). Existe ainda o critério de redução de volume, durante a digestão o volume de sólidos voláteis é habitualmente reduzido em cerca de 40% a 60% e uma quantidade de sobrenadante pode retornar para tratamento. Tendo em conta a variação do volume de sólidos no interior, é possível calcular o volume de digestor necessário a partir da seguinte fórmula (Qasim, 1999):

$$V = [Q_{in} - 2/3 (Q_{in} - Q_{out})] D_T \quad (\text{equação 1.3})$$

Onde:

V = Volume do digestor (m^3)

Q_{in} = Volume de lamas afluente (m^3/dia)

Q_{out} = Volume de lamas retiradas (m^3/dia)

D_T = Período de digestão (dias)

Sistema de Aquecimento

A escolha e manutenção da temperatura de operação são aspetos fundamentais para a eficiência do processo. (Malina e Pohland, 1992). Na tentativa de manter o parâmetro temperatura estável, em especial em casos de temperatura de operação elevada, é necessário instalar equipamentos de aquecimento no digestor, bem como sistemas de isolamento para evitar perdas de energia pelas paredes, chão e cobertura. O sistema mais comum de aquecimento, consiste em bombear as lamas e o sobrenadante para permutadores de calor externos e, em seguida, de volta para dentro do digestor (Metcalf & Eddy, 2003).

Condições de Mistura

A mistura da matéria dentro do digestor é essencial para se criar uniformidade física, química e biológica ao longo do digestor, de modo a prevenir a formação de sobrenadante e a deposição de substrato no fundo do reator (Turovskiy e Mathai, 2006). Outra vantagem adicional de se garantir uma mistura completa é que, em caso de entrada no digestor de produtos ou substâncias tóxicas no digestor, a sua dispersão é efetuada de forma mais imediata (Qasim, 1999). A partir da subida das bolhas de biogás até à superfície, e das correntes térmicas geradas dentro do digestor pelo equipamento de aquecimento, produz-se uma certa agitação natural. Não obstante, o equipamento auxiliar promotor de agitação é fundamental para se otimizar a performance do sistema e se garantir uma mistura uniforme de todo o substrato, que

de outra forma não seria possível (Turovskiy e Mathai, 2006). Os sistemas de agitação tipicamente mais usados têm por base um funcionamento por injeção de gás, agitação mecânica ou bombagem mecânica (Metcalf & Eddy, 2003).

2.3.5 Variáveis ambientais que Influenciam o Processo de Digestão Anaeróbia

Atualmente, os sistemas de digestão anaeróbia utilizados são considerados como uma tecnologia consolidada e que é utilizada em todo o mundo para tratamento de diversos tipos de resíduos e de águas residuais de diferentes setores (Donoso-Bravo *et al*, 2009). No entanto, existem diversos fatores ambientais que podem afetar o processo de digestão anaeróbia. Os mais importantes são a temperatura, o pH, a alcalinidade e a presença e respectivas concentrações de nutrientes e de substâncias tóxicas nos digestores anaeróbios (Metcalf & Eddy, 2003).

Temperatura

A temperatura de funcionamento do digestor é um dos fatores que mais influencia todo o processo de digestão anaeróbia. As taxas de crescimento dos microrganismos e de estabilização de sólidos aumentam e diminuem, consoante a temperatura de funcionamento do digestor (Qasim, 1999). Os digestores são projetados para funcionar frequentemente na gama de temperatura mesófila ou, com menor frequência, na gama termófila (Turovskiy e Mathai, 2006). Para a gama mesófila e termófila, as temperaturas ideais rondam os 35°C e os 54°C, respetivamente, dado ser á volta destes valores que as taxas de crescimento e de estabilização são maximizadas (Qasim, 1999). A temperatura é assim determinante para que os diferentes estágios da digestão anaeróbia se sucedam sem problemas, principalmente no caso da hidrólise e metanogénese, visto estas serem as fases em que as bactérias apresentam uma maior sensibilidade às condições ambientais (Metcalf & Eddy, 2003).

A temperatura de operação escolhida também tem grande influência no tempo mínimo de retenção de sólidos necessário para se atingir uma determinada redução na quantidade de sólidos voláteis existentes (Vesilind, 2003). Na Tabela 2.6 é possível verificar a relação entre estes parâmetros para reatores de alta carga a funcionar a temperaturas mesófilas. Quanto mais perto a temperatura estiver do valor ideal, mais depressa os organismos se multiplicam, logo a estabilização anaeróbia de lamas também se dá de modo mais rápido (Turovskiy e Mathai, 2006).

Tabela 2.6 – Relação entre a temperatura de operação e o tempo de retenção de sólidos em digestores anaeróbios de alta carga (adaptado de Turovskiy e Mathai, 2006)

Temperatura de operação (°C)	Tempo de retenção de sólidos mínimo recomendável (dias)
18	28
24	20
30	14
35	10
40	10

O estabelecimento da temperatura de operação é importante, mas mais importante é a manutenção desse valor, tendo em conta que as bactérias são sensíveis a alterações de temperatura. Variações de temperatura superiores a 1°C por dia podem resultar no decaimento da produção de biogás ou mesmo na inibição do processo de digestão anaeróbia. Um sistema de digestão bem projetado tem de ter capacidade para evitar mudanças de temperatura superiores a 0,5°C por dia (Vesilind, 2003).

pH e Alcalinidade

Cada comunidade de microrganismos possui diferentes capacidades de adaptação a diferentes intervalos de pH, no entanto na digestão anaeróbia, o controlo do pH é especialmente importante devido à sensibilidade das bactérias metanogénicas em relação a este parâmetro. Os digestores operam vulgarmente dentro dos limites de pH confortáveis para o desenrolar da atividade dos microrganismos metanogénicos, de forma a evitar desequilíbrios no sistema (Rajeshwari *et al*, 1999). O nível ótimo de pH para a produção de biogás situa-se entre os 6,8 e os 7,2 (Vesilind, 2003). Valores de pH inferiores a 6,0 conduzem à inibição da atividade das bactérias metanogénicas, resultando numa acumulação de ácidos orgânicos voláteis no reator e no fim da produção de biogás (Malina e Pohland, 1992). A produção de ácidos voláteis durante a acidogénese tem tendência para baixar os valores de pH, porém esta redução é normalmente contrariada pela produção de alcalinidade na forma de amónia e bicarbonato, durante a metanogénese (Turovskiy e Mathai, 2006). No entanto, o ponto-chave para o equilíbrio do pH é a relação entre as concentrações de dióxido de carbono e de bicarbonato. O dióxido de carbono ao solubilizar forma o ácido carbónico, o qual é o maior consumidor de alcalinidade no digestor. Concentrações demasiado elevadas de ácido carbónico, por vezes, requerem a adição externa de bicarbonato de sódio, de cal ou carbonato de sódio, de forma a aumentar o poder tampão do meio (Metcalf & Eddy, 2003).

Nutrientes

As comunidades bacterianas requerem macronutrientes (azoto, fósforo, cálcio, entre outros), micronutrientes (Cobre, Níquel, Ferro, entre outros) e quantidades que chegam a ser vestigiais

de outras substâncias para se reproduzirem de modo sustentado. Embora estes elementos sejam necessários em pequeníssimas quantidades, a falta destes nutrientes pode afetar bastante o desenvolvimento e performance dos microrganismos. A proporção ótima de C:N:P é de 100:2,5:0,5 para que a produção de biogás apresente bons valores. A concentração de nutrientes mínima necessária para garantir a atividade das bactérias pode ser calculada com base na concentração de CQO presente na água residual, no rendimento celular e concentração de nutrientes presente nas células das bactérias. No geral, a concentração de nutrientes no efluente deve ser duas vezes a concentração de nutrientes mínima necessária, de modo a que durante o processo de digestão anaeróbia este fator não seja limitativo (Rajeshwari *et al.*, 1999).

Substâncias Tóxicas

As substâncias tóxicas são comumente a causa apontada para problemas e falhas verificadas nos reatores anaeróbios, no entanto, a presença por si só destas substâncias não significa que o processo não possa continuar a ocorrer. Um elemento é considerado inibidor ao causar um efeito adverso no desenvolvimento e atividade das comunidades microbianas, quando presente em concentrações substanciais no digestor anaeróbio (Chen *et al.*, 2008).

Uma variedade de substâncias possui efeitos inibitórios a diferentes graus no processo de digestão anaeróbia (Lin Chen *et al.*, 2014). As consequências da sua presença podem ir desde a total inibição do processo (toxicidade aguda), a quebras na taxa de atividade dos microrganismos e da estabilização de lamas, passando pela redução da quantidade e qualidade de biogás produzido (Chen *et al.*, 2008). A origem destas substâncias pode depender do tipo e características de água residual afluyente à ETAR ou ser resultado das reações metabólicas das bactérias, ou seja, um subproduto do próprio processo de digestão. Nos elementos tóxicos ou inibitórios incluem-se compostos orgânicos, catiões de metais leves, amônia, sulfuretos, metais pesados, alguns materiais inorgânicos e os emergentes nano materiais. Estes tóxicos afetam principalmente as atividades das bactérias produtoras obrigatórias de H₂ e das bactérias metanogénicas. Contudo, globalmente, e devido às diferenças entre comunidades de microrganismos existentes e à variação da composição das águas residuais, os efeitos inibitórios registados na etapa de digestão anaeróbia têm sido bastante diferentes (Lin Chen *et al.*, 2014). Na Tabela 2.7 é possível observar o valor de concentrações inibitórias de alguns dos compostos mais frequentemente encontrados em processos anaeróbios.

Tabela 2.7 – Concentrações inibitórias de compostos presentes em processos anaeróbios (adaptado de Parkin e Owen, 1986 apud Metcalf & Eddy, 2003; Turovskiy e Mathai, 2006; Vesilind, 2003)

Substância	Concentração inibitória moderada (mg/L)	Concentração inibitória forte (mg/L)
Sódio (Na^+)	3 500 – 5 500	8 000
Potássio (K^+)	2 500 – 4 500	12 000
Cálcio (Ca^{2+})	2 500 – 4 500	8 000
Magnésio (Mg^{2+})	1 000 – 1 500	3 000
Amônia (NH_4^+)	1 500 – 3 000	3 000
Sulfureto (S^{2-})	200	200
Cobre (Cu^{2+})	-	0,5 (solúvel) 50 – 70 (total)
Crômio (Cr VI)	-	3,0 (solúvel) 200 – 250 (total)
Crômio (Cr III)	-	2,0 (solúvel) 180 – 420 (total)
Níquel (Ni^{2+})	-	30 (total)
Zinco (Zn^{2+})	-	1,0 (solúvel)

Efluentes com elevados teores de salinidade são especialmente difíceis de tratar pelo processo de digestão anaeróbia. Se a ETAR tratar águas ricas em sais, os valores de condutividade presentes no digestor anaeróbio serão mais elevados que o normal, o que pode indicar problemas na estabilização de lamas e produção de biogás (Zhang *et al.*, 2012). Tendo em conta a importância desta temática para um dos casos de estudo da atual dissertação, o subcapítulo 2.3.6 é dedicado a este assunto.

2.3.6 Influência da Salinidade no Funcionamento da Digestão Anaeróbia

Conceitos Básicos sobre Salinidade e Condutividade

A salinidade é um indicador da concentração de sais dissolvidos numa massa de água. Os principais iões responsáveis por esta propriedade da água são (UNEP GEMS/Water Programme, 2008):

- Os catiões como o Ca^{2+} , o Mg^{2+} , o Na^+ e o K^+ ;
- Os aniões como os carbonatos (CO_3^{2-} e HCO_3^-), o sulfato (SO_4^{2-}) e o cloreto (Cl^-).

A composição iónica das águas superficiais é relativamente estável e modifica-se pouco com os processos biológicos que ocorrem na massa de água (UNEP GEMS/Water Programme, 2008). Na Tabela 2.8 é possível observar a concentração dos iões principais na água do mar.

Tabela 2.8 – Concentração dos iões principais na superfície da água do mar (adaptado de Pilson, 2013)

Ião	Concentração (g de ião/Kg de água do mar)
Na ⁺	10,782
Mg ²⁺	1,284
Ca ²⁺	0,415
K ⁺	0,399
Cl ⁻	19,353
SO ₄ ²⁻	2,712
HCO ₃ ⁻	0,111
CO ₃ ²⁻	0,011

A salinidade pode ser medida por comparação entre a concentração de sólidos dissolvidos numa amostra de água face a uma solução padrão. Os sólidos dissolvidos podem ser estimados a partir da turvação ou da condutividade elétrica da água (UNEP GEMS/Water Programme, 2008). A turvação mede as propriedades de transmissão da luz de uma água e pode ser usada como indicador da qualidade das descargas de águas residuais ou das águas naturais, pois existe uma razoável relação entre a turvação e a concentração de sólidos totais suspensos. Os resultados da turvação são apresentados em unidades nefelométricas de turvação (NTU).

A condutividade elétrica da água é uma medida que classifica a capacidade da água para conduzir uma corrente elétrica. A corrente elétrica é transportada pelos iões presentes na massa de água, pelo que a condutividade aumenta quando a concentração de iões existente também aumenta (Metcalf & Eddy, 2003). A estimacão da condutividade é frequentemente usada como medição da salinidade existente numa massa de água, apresentando valores consideravelmente mais elevados em sistemas de águas salinas (UNEP GEMS/Water Programme, 2008). A condutividade elétrica é expressa em millisiemens por metro (mS/m) no Sistema Internacional, usando-se por vezes em estudos a medida milisiemens por centímetro (mS/cm) (Metcalf & Eddy, 2003). Em oceanografia utiliza-se a escala prática de salinidade (*Practical Salinity Scale 1978*), onde a salinidade é definida como sendo uma proporção entre o seu teor em sais e a sua condutividade elétrica, cuja medida adimensional é designada de Unidade Prática de Salinidade (PSU – *Practical Salinity Unit*) (UNESCO, 1985).

Águas residuais com teores de condutividade elevados

A presença de sais em águas residuais é bastante comum em determinadas circunstâncias, seja por motivos naturais ou por ações antropogénicas. É normal que o efluente bruto de uma ETAR - onde chegam águas residuais de processos industriais como os da indústria de transformação de peixe - contenha frequentemente uma grande quantidade de sais inorgânicos solúveis (Zhang *et al.*, 2012). As águas residuais que afluem a ETAR localizadas em zonas

costeiras e junto a estuários, também costumam apresentar teores em sais elevados, devido a intrusões de águas salinas e salobras nas redes de drenagem (Palmeiro-Sánchez *et al.*, 2013). Os estuários são ecossistemas de transição, influenciados por rios de água doce mas possuidores de águas de carácter parcialmente salgado, devido à sua proximidade a águas costeiras, por isso possuem valores de condutividade elétrica elevados (APA, 2015).

A ETAR de um dos casos de estudo desta dissertação situa-se junto ao estuário do Rio Tejo, sendo por isso afetada pela dinâmica do estuário. No ano de 2014 foi realizada, neste estuário, uma investigação, com esta temática, pelo que é necessário e útil rever e expor alguns dados deste estudo que ajudam a entender alguns conceitos estreitamente relacionados com o tema vertente. A investigação promovida e efetuada pela SIMARSUL em colaboração com a FCT/UNL, localizou-se no esteiro de Coina, junto à ETAR do Seixal. O estudo teve como objetivo identificar os principais impactes causados pela presença de sais acima do normal na água residual afluente ao sistema de tratamento da ETAR do Seixal. Foram apurados valores de condutividade no estuário e nos caudais afluentes à ETAR tendo em vista a verificação de intrusões salinas nesta instalação. As medições realizadas apontam para valores entre os 40 mS/cm e 50 mS/cm de condutividade existente no estuário do Tejo. Efetuaram-se também medições no efluente bruto que aflui à ETAR de Fernão Ferro e na água de abastecimento à população do Seixal. Na ETAR de Fernão de Ferro não existia qualquer indício de intrusões salinas, o que se veio a confirmar através dos valores baixos de condutividade registados na ordem dos 0,9 mS/cm e 1,6 mS/cm. As medições de condutividade na água de abastecimento apresentam valores entre os 0,3 mS/cm e os 0,5 mS/cm. Com base nas medições realizadas neste estudo, considera-se que existem intrusões salinas na rede de água residual afluente a uma ETAR, se os valores de condutividade medidos no interior da instalação forem superiores a 2 mS/cm (Serrano, 2014). Valores de condutividade desta ordem podem indicar problemas no funcionamento dos equipamentos e processos instalados em uma ETAR, nomeadamente nos processos anaeróbios.

Problemas no Funcionamento dos Digestores Anaeróbios

Os efeitos da presença de águas de características salobras na performance do processo de digestão anaeróbia têm sido pouco estudados, no entanto, é possível constatar dificuldades no tratamento deste tipo de águas e alguns efeitos dos principais constituintes destas águas no funcionamento do digestor anaeróbio (Lefebvre e Moletta, 2006; Palmeiro-Sánchez *et al.*, 2013). Águas residuais com teores em salinidade elevados são sempre difíceis de tratar, através do recurso a tecnologias anaeróbias comuns como a digestão anaeróbia (Zhang, *et al.*, 2012). A presença de sais em concentrações acima do normal no digestor anaeróbio pode inibir o funcionamento eficaz deste processo, caso os microrganismos não estejam aclimatados a esta situação (Chen *et al.*, 2008; Jeison *et al.*, 2007). Existem registos em que a atividade dos microrganismos presentes em reatores anaeróbios é afetada a partir das 2 500 mg/L de NaCl e

fortemente inibida a partir de concentrações que variam entre as 7 000 mg/L e 10 000 mg/L de Na, sendo estes os sais normalmente mais comuns de se encontrar no efluente (Jeison et al. 2008; Lefebvre e Moletta, 2006; Vallero et al., 2003). As comunidades bacterianas podem entrar em *stress* osmótico gerando a plasmólise das suas células por perda de água por osmose, o que conduz à lise celular e consequentes problemas na atividade enzimática das bactérias (Vallero et al., 2003; Zhang *et al.*, 2013).

Os catiões e aniões em solução necessitam de estar sempre associados, no entanto, a toxicidade dos sais é predominantemente determinada pelos catiões (McCarty e McKinney, 1961 *apud* Chen *et al.*, 2008). Em seguida, serão descritos os efeitos na digestão anaeróbia de alguns dos principais componentes presentes em águas salobras, tendo em consideração que os efeitos registados podem ser minorizados ou amplificados, perante situações de antagonismo ou sinergismo entre catiões e aniões. No entanto, dada a amplitude e diversidade de pequenas alterações que podem ocorrer, derivadas de variáveis relações de antagonismo e sinergismo que se podem encontrar na literatura, estas relações não serão relatadas.

Sódio e Cloro

O Na e o Cl são correspondentemente, o catião e o anião mais presentes em águas residuais influenciadas por efluentes salinos, e quando se associam formam o cloreto de sódio (NaCl) (Feijoo *et al.*, 1995). As comunidades de bactérias podem ser afetadas por vários sais, mas são principalmente inibidas quando a concentração de cloreto de sódio no meio é elevada, sendo que o Na é o principal responsável pela toxicidade desta substância (Lefebvre e Moletta, 2006). O efeito do Na na digestão anaeróbia, sobretudo na metanogénese tem sido particularmente estudado. O Na em concentrações reduzidas é essencial à metanogénese, valores de concentração entre 100 e 400 mg/L têm um efeito positivo no processo da digestão anaeróbia (Palmeiro-Sanchez *et al.*, 2013). A partir de concentrações de NaCl acima das 2 500 mg/L, o desempenho metabólico das bactérias metanogénicas começa a alterar-se. (Vallero et al., 2003). Decorrente da observação da Tabela 2.7, é possível verificar que, a partir de concentrações de Na superiores a 3 500 mg/L existe inibição do processo de digestão anaeróbia. Na restante literatura consultada, é possível verificar a existência de bastantes relatos em que a atividade da bactéria metanogénica cessa a partir de concentrações de Na superiores a 10 000 mg/L, principalmente em digestores não aclimatizados a efluentes salinos.

Magnésio

Os efeitos do catião Mg na digestão anaeróbia não têm sido alvo de muita investigação científica, contudo, sabe-se que concentrações até as 400 mg/L são benéficas para o desenvolvimento dos microrganismos (Schmidt *et al.*, 1992). Concentrações de Mg acima das

1 000 mg/L podem possuir carácter inibitório na atividade bacteriana, conforme estabelecido na Tabela 2.7.

Cálcio

O Ca é conhecido por ser essencial para o crescimento de certas estirpes de bactérias metanogénicas (Murray e Zinder, 1985). Em quantidades excessivas leva à precipitação de carbonato e fosfato, o que leva à redução de atividade dessas mesmas bactérias, bem como à perda de poder tampão e de nutrientes (Van Langerak *et al*, 1998). A literatura analisada não é totalmente consensual, mas pode-se indicar uma concentração de 150 a 200 mg/L como a ideal, enquanto concentrações de 2500 a 4500 mg/L e 8000 mg/L inibem respetivamente de forma moderada e de forma forte o processo de digestão anaeróbia como se pode observar igualmente na Tabela 2.7.

Potássio

No digestor anaeróbio, a presença de K não costuma gerar problemas, dada a sua concentração ser geralmente reduzida. Os efeitos tóxicos inclusivamente desta substância no processo de digestão anaeróbia têm sido pouco investigados (Chen *et al.*, 2008; Zhang *et al.*, 2013). Concentrações abaixo das 400 mg/L causam um efeito positivo na performance do digestor, podendo ser uma vantagem para o bom funcionamento do processo. Todavia concentrações elevadas de potássio causam um efeito inibitório principalmente na digestão em gamas de temperatura termofila (Chen *et al.*, 2008). Na Tabela 2.7 é possível observar as concentrações de K que geram inibição no processo de digestão anaeróbia.

Sulfato

O sulfato é um constituinte comum de muitas águas residuais e em reatores anaeróbios é reduzido pelas bactérias a sulfureto (O'Flaherty *et al*, 1998). Esta redução é efetuada por dois grupos principais de bactérias, um que reduz compostos como o ácido láctico a acetato e CO₂ e outro que converte o acetato completamente a CO₂ e bicarbonato. A redução de sulfato a sulfureto é responsável por inibição a dois níveis. O primeiro acontece devido à competição entre bactérias por substratos orgânicos e inorgânicos, o que conduz a uma diminuição na produção de biogás. O segundo nível de inibição resulta do facto de o sulfureto ser tóxico para grupos variados de bactérias (Colleran *et al*, 1995; Colleran *et al*, 1998). A partir do momento em que o sulfureto apresenta uma concentração de 200 mg/L é passível de causar inibição moderada ou forte ao processo de digestão anaeróbia, como se pode observar na Tabela 2.7. Este valor limite pode ser um pouco diferente consoante as condições de funcionamento do digestor, podendo ser inferior a 200 mg/L face a um pH superior a 7,2 (O'Flaherty *et al*, 1998).

As informações relativas à influência do pH nas concentrações de sulfureto são raras na literatura, o que impossibilita retirar conclusões seguras nesta temática (Chen *et al*, 2008).

Adaptação dos microrganismos a efluentes salobros

Os microrganismos possuem a capacidade de se adaptar a concentrações de sais acima do normal, mas esta capacidade de adaptação depende fortemente do tipo de cultura e da composição da água residual. (Oh *et al*, 2008). Além de que uma adaptação com sucesso apenas faz elevar um pouco a concentração limite de inibição a que os organismos podem ser sujeitos, o que em casos extremos não resolve o problema (Zhang *et al.*, 2012) Nos casos em que o caudal afluente ao processo de digestão anaeróbia possui concentrações de sais demasiado elevadas, a melhor opção a tomar deverá ser a aplicação complementar de tecnologias físico-químicas de remoção de sais na linha de tratamento de uma ETAR.

2.3.7 Produtos da Digestão Anaeróbia

Os principais produtos do processo de digestão anaeróbia são o biogás e as lamas digeridas, sendo que na digestão em baixa carga e em dois estágios também é preciso lidar com a produção de sobrenadante (Jenicek *et al* 2014). O sobrenadante gerado num sistema anaeróbio costuma conter elevadas concentrações de sólidos dissolvidos e suspensos, matéria orgânica, nutrientes e outras substâncias em menor quantidade (Turovskiy e Mathai, 2006). O sobrenadante é usualmente enviado para o início da estação de tratamento, o que pode significar um aumento considerável na carga a tratar (Qasim, 1999).

As lamas digeridas possuem uma concentração entre 4% a 6% de matéria sólida seca e após estabilização são habitualmente enviadas para desidratação. (Qasim, 1999). A eliminação de lamas gera custos consideráveis, por isso a reutilização destas tem sido encorajada por motivos de sustentabilidade (Jenicek *et al* 2014). Contudo encontrar um destino final para as lamas, por vezes, pode ser um problema, pois estas para serem utilizadas em certas atividades como por exemplo na agricultura, necessitam de cumprir requisitos legislativos e de serem aceites pelos agricultores, o que nem sempre acontece (Vesilind, 2003).

Uma digestão anaeróbia bem conseguida produz biogás formado por 65% a 70% de CH₄, por 25% a 30% de CO₂ e por pequenas quantidades de N, ácido sulfídrico e H₂ (Metcalf & Eddy, 2003). O poder calorífico do biogás produzido encontra-se normalmente entre os 21 000 KJ/m³ a 25 000 KJ/m³ (Qasim, 1999). A quantidade de biogás gerado pode ser estimada a partir de um coeficiente de 0,8 m³/Kg a 1,1 m³/Kg de sólidos voláteis destruídos durante a digestão (Vesilind, 2003). O valor de biogás produzido será próximo do seu máximo, se o digestor tiver uma percentagem elevada de matérias gordas e desde que o tempo de retenção de sólidos seja adequado à metabolização lenta destas substâncias. Numa ETAR de média a grande

dimensão que possua um digestor anaeróbio a operar bem, frequentemente, este irá produzir uma quantidade de biogás capaz de gerar mais energia que a necessária, para colmatar os gastos energéticos do seu funcionamento. A quantidade extra de energia concebida pode ser usada na satisfação de outras necessidades energéticas das instalações da ETAR ou pode ser vendida a outra concessionária local (Turovskiy e Mathai, 2006). O sistema de aproveitamento de biogás que converte este a energia útil inclui no seu inventário, captadores de biogás, tubagens para a condução do biogás, válvulas redutoras de pressão, queimadores de biogás, compressores de biogás e gasómetros para o seu armazenamento (Qasim, 1999). O biogás ainda possui a particularidade de poder ser utilizado em sistemas de cogeração, os quais geram energia elétrica e térmica que em ambos os casos pode ser aproveitada. Na cogeração, o biogás é usado como fonte de energia de um gerador que produz energia elétrica; a água de arrefecimento descarregada a temperaturas na ordem dos 70°C a 82°C e o biogás de escape provenientes do engenho são utilizados como energia térmica (Turovskiy e Mathai, 2006).

2.3.8 Modelação: Principais Tipologias Aplicadas

No processo de digestão anaeróbia é crítico garantir o funcionamento seguro de todo o sistema. Os digestores são habitualmente concebidos com base numa condição de operação nominal relativamente estável, que na prática raramente é cumprida. O processo está sujeito a muitas variações na quantidade e qualidade de carga a tratar, o que pode resultar na redução do desempenho do digestor ou mesmo na sua falha total (Holubar *et al.*, 2001). A digestão anaeróbia pela sua instabilidade, é um processo ainda alvo de investigação e em contínuo desenvolvimento (Lauwers *et al.*, 2013). Desde finais da década de sessenta que se desenvolvem modelos matemáticos, com o intuito de aumentar o conhecimento científico acerca da digestão anaeróbia. Os modelos criados com o avançar do tempo tornaram-se cada vez mais complexos e explicativos (Yu *et al.*, 2013). Existe um sentimento geral na comunidade científica, de que o desenvolvimento destes modelos é de grande contribuição para um melhor entendimento da dinâmica do processo de digestão (Lauwers *et al.*, 2013).

No geral, quando se opta por aplicar um processo de modelação à digestão anaeróbia, é pelos seguintes objetivos: avaliar ou compreender o comportamento do sistema e as interações entre os seus componentes; prever o comportamento do sistema no futuro ou sob determinadas circunstâncias; otimizar o funcionamento da digestão anaeróbia. Todo o processo de modelação (a estrutura do modelo, a identificação dos valores dos parâmetros selecionados e a planificação de medições experimentais) deve estar em consonância com os objetivos estabelecidos (Donoso-Bravo *et al.*, 2011).

A classificação de modelos é realizada com base em duas categorias, os modelos dinâmicos e os modelos não dinâmicos. Os modelos dinâmicos são capazes de executar previsões contínuas ao longo do tempo ou, pelo menos, previsões em intervalos discretos. Os modelos

não dinâmicos conseguem apenas concretizar previsões para variáveis independentes do tempo. A chave dos modelos não dinâmicos costuma residir nas ligações estequiométricas entre os tipos de substrato e os produtos, sendo geralmente utilizadas equações de cálculo com os elementos carbono, hidrogénio, oxigénio e azoto. No entanto, esta abordagem raramente é aplicada, porque as equações estequiométricas não contemplam a matéria não biodegradável nem o crescimento da biomassa ao longo do processo (Lauwers *et al.*, 2013). Em relação aos modelos mais utilizados, os dinâmicos, é possível observar na Figura 2.4, em esquema, as várias etapas da sua base processual.

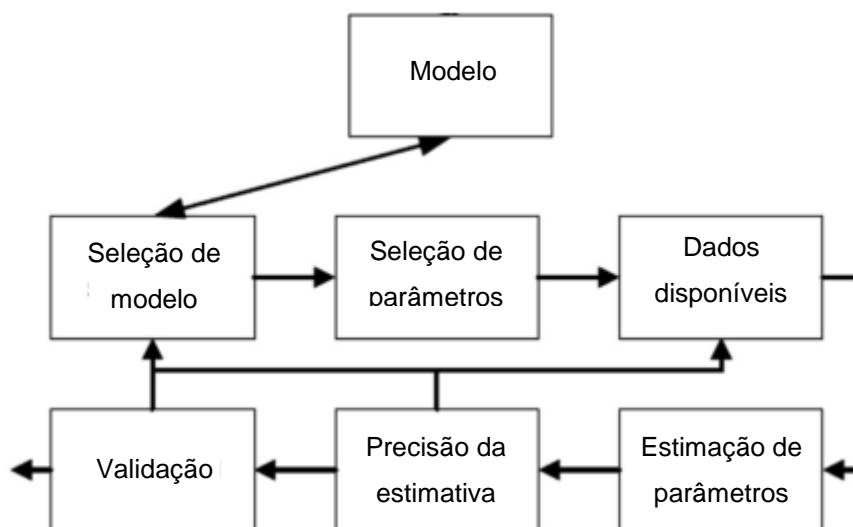


Figura 2.4 – Procedimento esquemático dos modelos dinâmicos (adaptado de Donoso-Bravo *et al.*, 2011; Lauwers *et al.*, 2013)

Quando se decide por um modelo dinâmico, aquando da seleção de modelo, pode-se optar entre modelos matemáticos “mecanizados” ou modelos orientados por dados. Os modelos mecanizados podem ser principalmente de dois tipos, dedutíveis dado que usam a informação disponível para descrever as reações bioquímicas que ocorrem durante a digestão, ou modelos em que os parâmetros têm uma interpretação física mas ajustável. Estes últimos têm sido dos mais aplicados em modelação do processo de digestão anaeróbia (Lauwers *et al.*, 2013).

Os primeiros modelos matemáticos de inspiração mecânica criados remontam à década de 70 e eram baseados em uma única etapa limitante, a qual dependia de inúmeras condições como as características das águas residuais, tempo de retenção de sólidos ou a temperatura do digestor. Alguns destes modelos consideravam a metanogénese como a etapa limitante, ao passo que outros consideravam a conversão de ácidos gordos ou a hidrólise (Appels *et al.*, 2008). Com o avançar do tempo e o desenvolvimento dos estudos microbiológicos, foram criados modelos capazes de incorporar mais pormenores e reações cinéticas mais detalhadas (Donoso-Bravo *et al.*, 2011). A variedade de modelos desenvolvidos levou à conclusão que era necessário criar uma solução que convergisse e consolidasse as várias opções de modelação disponíveis. Em resposta a esta necessidade é criado em 1997 pelo “IWA Task Group on

Mathematical Modelling of Anaerobic Digestion Processes” o *Anaerobic Digestion Model No. 1* (ADM1), modelo que vem unificar a base da modelação relativa à digestão anaeróbia (Lauwers *et al.*, 2013). A estrutura do modelo inclui múltiplas etapas descritivas dos processos bioquímicos e físico-químicos da digestão anaeróbia. Implementado como diferencial e conjunto de equações algébricas, o modelo contém 26 variáveis de concentração de estado dinâmicas e oito variáveis algébricas implícitas por reator. Se pelo contrário, estiver implementado apenas como um diferencial de equações, o modelo possui 32 variáveis de concentração de estado dinâmicas (Batstone *et al.*, 2002). O ADM1 tem sido, sem dúvida, um dos modelos mais utilizados na modelação do processo de digestão anaeróbia, sendo possível encontrar diversos estudos científicos onde este modelo é empregue (Blumensaat e Keller, 2005; Derbal *et al.*, 2009; Lee *et al.*, 2009).

Os modelos dinâmicos até agora discutidos são todos eles baseados em equações que expressam um balanço de massas. Todavia existem também os modelos dinâmicos orientados por dados, mais conhecidos a nível global por modelos “*black-box*”. Este tipo de modelo relaciona variáveis de entrada com variáveis de saída, sem possuir qualquer conhecimento prévio das reações físicas e químicas que ocorrem no processo alvo de modelação (Holubar *et al.*, 2001; Lauwers *et al.*, 2013; Worden *et al.*, 2005; Yu *et al.*, 2013). Dentro do tipo de modelos dinâmicos orientados por dados que se podem usar para modelação da digestão anaeróbia, estão as RNA (Holubar *et al.*, 2001; Lauwers *et al.*, 2013; Worden *et al.*, 2005; Yu *et al.*, 2013). O interesse nesta tecnologia de modelação advém da sua capacidade de aprendizagem, da sua adaptabilidade, da capacidade de generalização que possui, da sua tolerância a erros e do facto de poder processar informação de modo linear ou não linear (Güçlü *et al.*, 2011; Haykin, 1999). As RNA têm-se mostrado bastante competentes na modelação de sistemas de energia multidimensionais e não-lineares, quando são necessárias previsões precisas de certos parâmetros (Nikpey *et al.*, 2013). Na digestão anaeróbia, as RNA podem ser usadas para fins de predição como a previsão dos valores de biogás, das concentrações de elementos vestigiais, entre outras ou simplesmente para controlo do processo de digestão. Nos próximos anos espera-se que se continuem a registar novos avanços nestes modelos, tal como na restante tipologia de modelação aplicada ao processo de digestão anaeróbia, tendo-se em especial atenção o aspeto biológico, ou seja, a relação entre a performance do digestor e as comunidades de bactérias (Lauwers *et al.*, 2013).

Por fim, é preciso realçar que para a construção e validação dos modelos mencionados, a monitorização e controlo do funcionamento dos digestores anaeróbios são absolutamente vitais. Só será possível continuar a progredir nesta área, se pelo menos os parâmetros principais da digestão anaeróbia forem monitorizados de forma contínua ou com uma frequência muito elevada (Adgar *et al.*, 2000; Appels *et al.*, 2008).

2.4 Redes Neurais Artificiais

As RNA são modelos computacionais inspirados na biologia do cérebro humano, adaptadas para processar informação geralmente de forma não linear (Ding *et al.*, 2011). As RNA são muito menos complexas que o cérebro humano, no entanto, o seu funcionamento é similar ao do cérebro na memorização de certos padrões e no reconhecimento de novos padrões semelhantes a outros já observados. O primeiro passo para a compreensão deste tipo de modelo computacional, é a realização de uma análise ao funcionamento do cérebro humano, mais concretamente dos neurónios celulares (Kasabov, 1996; Lingireddy e Brion, 2005).

2.4.1 Funcionamento do Neurónio Celular

O cérebro humano pode ser descrito como uma rede neuronal biológica, formada por um conjunto de neurónios interconectados entre si, os quais recebem e emitem elaborados padrões de sinais eléctricos (Rojas, 1996; Shiffman, 2012). Na Figura 2.5 é possível observar a estrutura simplificada do neurónio celular. Os ramos da esquerda, denominados dendrites, são os canais transmissores dos sinais que chegam ao neurónio. Os sinais atravessam o corpo da célula e são conduzidos até outro neurónio pelos axónios. A região onde um sinal é transferido do axónio de um neurónio para a dendrite de outro neurónio, é chamada de sinapse (Lingireddy e Brion, 2005; Rojas, 1996).

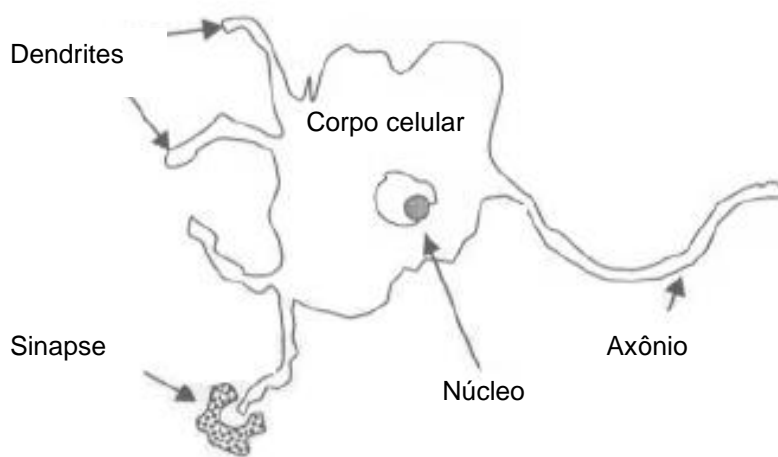


Figura 2.5 – Estrutura simplificada do neurónio celular (adaptado de Lingireddy e Brion, 2005)

Na Figura 2.6 é possível observar-se a forma como flui a informação de um neurónio para outro na biologia celular, face à representação esquemática do fluxo de informação do ponto de vista de um modelo matemático. As setas apontadas às dendrites (imagem da esquerda) ou aos nós mais pequenos (imagem da direita) representam os sinais que chegam aos neurónios 1, 2 e 3. No caso da imagem à esquerda, os sinais são trespassados de um axónio para uma dendrite nas sinapses, enquanto na imagem à direita, existe um nó intermédio entre dois neurónios (Lingireddy e Brion, 2005).

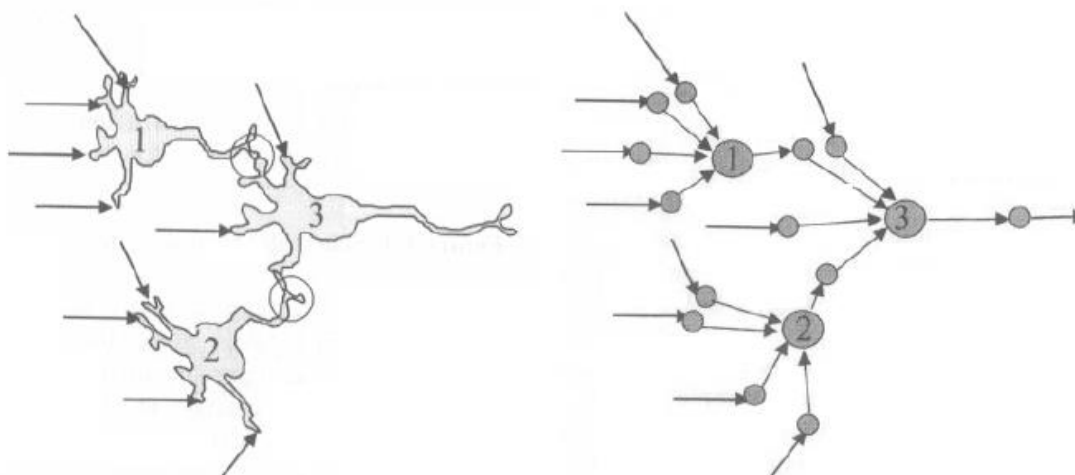


Figura 2.6 – Fluxo de informação na biologia celular face à representação esquemática desse fluxo em um modelo matemático (adpatado de Lingireddy e Brion, 2005)

Os sinais ou impulsos nervosos carregam informação que só pode ser processada através da interação de inúmeros neurónios e fibras nervosas (Kasabov, 1998). A força com que cada impulso é transmitido pode ser controlado através da alteração do tamanho da sinapse (Lingireddy e Brion, 2005). A existência de muitas ligações entre os neurónios determinam um alto nível de paralelismo, o que é característico dos mecanismos cerebrais. A habilidade do cérebro para analisar problemas complexos e reagir adequadamente a novas questões tem por base a sua capacidade heurística de tomar decisões com base em conhecimentos armazenados e de se adaptar a novas situações. O cérebro possui a capacidade de aprender e generalizar. (Kasabov, 1998). A capacidade de aprendizagem e processamento de informação das RNA apresenta características semelhantes à do cérebro, todavia o potencial e complexidade destas é bastante inferior (Martins, 1997).

2.4.2 Evolução histórica das Redes Neurais Artificiais

No ano de 1943, o neurocientista Warren S. McCulloch e o lógico Walter Pitts desenvolveram o primeiro modelo conceptual de uma RNA. Eles descrevem o conceito de neurónio como uma única célula viva dentro de uma rede de células que recebe informação, processa-a e gera uma resposta (Shiffman, 2012). Em 1949, o psicólogo Donald Hebb destacou a importância das conexões entre sinapses no processo de aprendizagem. A partir das ideias de Hebb, McCulloch e Pitts no ano de 1958 é descrito por Rosenblatt o primeiro modelo operacional de uma rede neuronal, denominado *Perceptron*. (Kasabov, 1998). Em 1969 Minsky e Papert provam que o modelo criado por Rosenblatt não consegue processar partes de funções lógicas simples, o que demonstrou os limites teóricos do modelo (Lingireddy e Brion, 2005; Kasabov, 1996). A demonstração da incapacidade do *Perceptron* resultou na desistência de muitos investigadores de prosseguirem os seus estudos no âmbito das redes neuronais. As RNA reemergiram após se terem registado alguns avanços teóricos no início da década de oitenta

(Lingireddy e Brion, 2005). Desde essa década que se voltou a apostar bastante no desenvolvimento das RNA, o que levou a que no presente existam muitos tipos de redes neuronais, as quais são aplicadas em diversas áreas (Ding *et al.*, 2011).

2.4.3 Funcionamento das Redes Neuronais Artificiais

O Neurónio Artificial

As RNA são sistemas computacionais conectores que processam a informação coletivamente e paralelamente (Shiffman, 2012). Este tipo de sistema é formado por um determinado número de nós, intitulados de neurónios, os quais se interligam entre si e são ordenados em diferentes camadas (Ding *et al.*, 2011; Kasabov, 1996; Lauwers *et al.*, 2013; Lingireddy e Brion, 2005). Os neurónios são auto adaptáveis, auto-organizáveis e possuem uma capacidade de aprendizagem em tempo real (Ding *et al.*, 2011). São capazes de ler a informação que recebem, processá-la e, consoante o género de informação recebida, emitem uma resposta (Shiffman, 2012). Na Figura 2.7 apresenta-se um modelo de um neurónio, o qual é baseado nos seguintes parâmetros (Holubar *et al.*, 2001; Kasabov, 1996):

- Dados de entrada (x_1, x_2, \dots, x_n), os quais possuem pesos sinápticos associados (w_1, w_2, \dots, w_n) onde está codificado o conhecimento da rede neuronal;
- Uma função (f) que calcula os sinais ou valores agregados dos dados de entrada no neurónio ($U = f(x, w)$);
- Um sinal de ativação da função s que calcula o nível de ativação do neurónio ($a = s(u)$);
- Uma função de saída que calcula o sinal ou valor de saída emitido pelo neurónio ($o = g(a)$), o valor de saída normalmente assumido é igual ao nível de ativação do neurónio, isto é, ($o = a$);
- De acordo com o tipo de valores de cada parâmetro acima pode tomar, os valores de entrada e saída no neurónio podem ser binários $\{0, 1\}$; bivalentes $\{-1, 1\}$; contínuos $[0, 1]$ ou números discretos de um determinado intervalo definido.

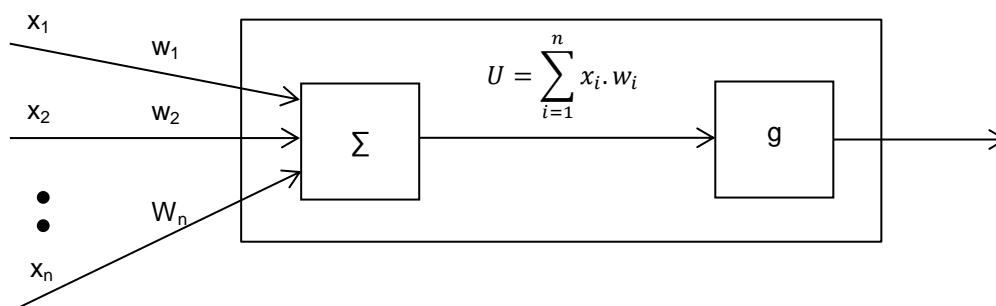


Figura 2.7 – Modelo de um neurónio artificial (adaptado de Kasabov, 1996)

Em suma, a ativação interna do sinal de saída é uma soma ponderada das suas entradas, no entanto a função de que define o nível de ativação também é usada na determinação do valor

final. Quando o sinal de saída é igual a 1, significa que o nível de ativação foi igualado ou excedido. Por outro lado, se o sinal de saída foi igual a 0, significa que a que o nível de ativação do neurónio não foi satisfeito (Rao e Rao, 1995).

Fases de Desenvolvimento de uma Rede Neuronal Artificial

Numa fase inicial é necessário o utilizador preparar a informação ou os dados que irá disponibilizar à rede. A qualidade da informação e a forma como esta é preparada acaba, muitas vezes, por condicionar os resultados finais gerados pelas RNA. Os métodos de preparação dependem muito do utilizador e dos objetivos que se pretendem atingir. Após a preparação de dados, a metodologia normalmente empregue no desenvolvimento das RNA é fundamentada em duas ou três fases.

A primeira fase diz respeito ao treinamento da rede, a segunda consiste em um período de testes e, caso se justifique, pode ainda existir uma fase de validação. Durante o treino ocorre a aprendizagem da rede, a partir de um conjunto de dados que lhe é fornecido e que exemplifica o resultado final pretendido. Os testes realizados permitem saber se os pesos sinápticos utilizados foram os mais corretos para a aprendizagem da rede. Se os resultados não forem bons, os pesos são novamente ajustados e o treino deve continuar. Quando os resultados são aceitáveis, o modelo pode passar à fase de validação, que confirma ou não a generalização do modelo. (Lingireddy e Brion, 2005).

A fase de treinamento é uma das mais importantes características das RNA e a que se dá bastante atenção (Vemuri, 1992). Uma rede neuronal pode ser treinada, de modo a que aprenda a classificar padrões, a agrupar dados, a avaliar uma função ou ainda tendo em vista outros objetivos. Esta etapa baseia-se sobretudo na alteração dos pesos sinápticos (Rao e Rao, 1995). Todo o conhecimento da rede está codificado nos pesos sinápticos, os quais são determinados no processo de aprendizagem da rede (Vemuri, 1992). Em relação a estes pesos associados aos neurónios é possível tomar duas ações (Rao e Rao, 1995):

- Começar com um determinado conjunto de pesos sinápticos e correr a rede neuronal (não existe treinamento da rede);
- Começar com um determinado conjunto inicial de pesos sinápticos, correr a rede neuronal permitindo que esta durante a sua aprendizagem vá modificando alguns ou todos os pesos definidos, de forma a se produzir um determinado sinal de saída desejado (existe treinamento).

Existem diferentes técnicas de treinamento, sendo que as mais comuns são o treino supervisionado, o treino não supervisionado e o treino reforçado (Vemuri, 1992). As redes podem ser treinadas de diferentes formas, porque em cada modelo as redes possuem um procedimento de aprendizagem incorporado e adaptável ao modelo de treino escolhido. O

procedimento responsável por permitir a aprendizagem da rede é chamado de algoritmo de aprendizagem. O algoritmo tem por base uma função, a qual estipula a lei de aprendizagem do modelo, dado ser responsável por alterar os pesos sinápticos de maneira ordenada, tendo em vista o cumprimento dos objetivos estipulados (Ding *et al.*, 2011; Haykin, 1999).

Em relação ao modelo de treino, este é considerado como supervisionado, quando se utilizam critérios externos que exemplificam à rede neuronal qual o resultado final pretendido (Vemuri, 1992). A rede é modelada de acordo com os tipos de estímulo fornecidos e as respostas que se esperam para cada tipo de estímulos. Durante a aprendizagem da rede, os parâmetros como os pesos sinápticos são ajustados sob influência da informação concedida à rede pelo utilizador (Haykin, 1999). As redes de treino supervisionado são um pouco mais simples de conceber e entender, face às redes treinadas de outra forma (Rao e Rao, 1995).

As redes de treino não supervisionado, ao contrário das supervisionadas, não são influenciadas por critérios externos ou, por outras palavras, não é dado nenhum exemplo à rede dos resultados pretendidos (Haykin, 1999; Rao e Rao, 1995). São fornecidos estímulos à rede, mas não as respostas desejadas, permitindo às redes se auto-organizarem, o que conduz a uma maior conexão entre os neurónios (Rao e Rao, 1995). Neste caso, a rede durante o treino tenta desenvolver modelos internos que lhe permitam aprender e reconhecer os padrões existentes nos estímulos ou dados fornecidos. (Vemuri, 1992). Neste estilo de treino, é muito comum falar-se também da aprendizagem competitiva, dado que para o treino não supervisionado ter sucesso, costuma ser necessário recorrer a este tipo de aprendizagem. Neste estilo de aprendizagem, os neurónios da última camada (camada de saída) competem entre si para serem ativados, ou seja, só é permitido a um neurónio emitir respostas aos estímulos fornecidos. Esta característica da aprendizagem competitiva torna este estilo bastante adequado para o treinamento não supervisionado de redes, dado os neurónios possuírem, neste caso, atributos propícios para a descoberta dos padrões estabelecidos na informação concedida (Haykin, 1999).

No treino reforçado, a aprendizagem da rede tem características do estilo de aprendizagem das redes de treino supervisionado e das redes de treino não supervisionado. No treino de reforço, não são fornecidos estímulos que demonstram à rede o resultado pretendido, mas no entanto concede-se informação à rede sobre o que é ou não um bom resultado, ou seja, em vez de informação instrutiva é concedida informação avaliativa (Vemuri, 1992). O processo de aprendizagem vai acontecendo à medida que as redes vão sabendo se estão ou não a produzir bons resultados, por outras palavras, à medida que recebem ou não um reforço positivo (Haykin, 1999).

A segunda fase de desenvolvimento de uma RNA corresponde à etapa de testes, nesta fase através de um número de testes executados aos estímulos concedidos à rede, que podem ou

não ter sido utilizados durante o treino, é possível verificar se o treinamento a que a rede foi sujeita detém potencial ou não de produzir bons resultados finais. (Taylor, 1996). Os testes realizados permitem estimar a qualidade da generalização realizada pela rede, ou seja, analisam se a rede está a responder a cada estímulo da forma pretendida (Kasabov, 1998). Caso os exemplos ou dados fornecidos para o treino da rede sejam em número insuficiente, a RNA não conseguirá produzir resultados precisos ou não conseguirá executar uma boa generalização. Se os resultados dos treinos forem maus, pode-se considerar que os dados providos não são suficientes. Quando os resultados do treino são bons e os resultados dos testes são maus, isto indica que a RNA apenas conseguiu memorizar os exemplos facultados, em vez de gerar uma boa generalização (Lingireddy e Brion, 2005).

A validação de resultados pode ou não ser realizada no final da fase de treino e de testes. A partir da validação de resultados pretende-se tomar conhecimento do quão bom são realmente os resultados produzidos por cada RNA. Nesta etapa, os resultados gerados são comumente comparados com resultados alcançados por outros modelos ou utilizadores, de forma a se comprovar quais são as RNA que valem a pena serem usadas na concretização dos objetivos estabelecidos (Kasabov, 1996; Lingireddy e Brion, 2005).

Concluída as fases de desenvolvimento de uma rede neuronal (fase de treino, de teste e hipoteticamente de validação), é possível utilizar estas em etapas de previsão, sendo muitas vezes este o objetivo que se pretende cumprir após a produção de RNA. A previsão realizada pelas RNA é o processo de gerar resultados para possíveis desenvolvimentos de um processo, com base em dados antigos e no desenvolvimento atual do processo (Kasabov, 1998). Há que realçar que a predição de valores por parte das RNA é uma característica bastante interessante destes modelos. No entanto, a realização desta etapa só é justificável se os resultados gerados pelas fases de treino e teste durante o desenvolvimento das RNA forem excelentes, sendo ainda aconselhado muitas vezes na literatura consultada, a execução da fase correspondente à validação de resultados.

2.4.4 Tipos de Redes Neurais Artificiais

Uma rede neuronal é uma estrutura computacional inspirada pelo estudo da biologia neuronal. Existem muitos tipos de diferentes RNA, desde as de estrutura simples até às de estrutura complexa, tal como existem muitas teorias sobre como se processa a informação nos cérebros humanos (Rao e Rao, 1995). A organização topológica dos neurónios e a forma como a estrutura da rede é estabelecida costuma-se designar por arquitetura de rede (Vemuri, 1992). Em relação a esta temática é preciso realçar que cada arquitetura de rede está intimamente conectada a um tipo de algoritmo de aprendizagem, responsável pelo treino de cada rede. No geral, podem ser identificados três tipos de classes principais de arquitetura de rede: as redes

de uma camada *feedforward*; as redes de multicamada *feedforward* e as redes neuronais recorrentes (Haykin, 1999).

Redes de Uma Camada *Feedforward*

Uma rede neuronal de uma camada é a forma mais simples de organização do tipo de redes possuidoras de uma arquitetura por camadas. Neste modelo, existe uma camada de entrada e uma camada de saída, sendo que a transmissão de sinais ocorre apenas no sentido dos neurónios da camada de entrada para os neurónios. Os neurónios da camada de saída são os únicos neste tipo de rede que possuem capacidade computacional para processar a informação, por isso este modelo é designado por rede de uma camada (Haykin, 1999). A maior limitação desta arquitetura é o facto de só conseguir resolver funções linearmente separáveis. O *Perceptron* desenvolvido por Rosenblatt no ano de 1958 faz parte desta categoria arquitetónica e foi a referida limitação que desencorajou alguns investigadores no passado a continuarem a desenvolver as RNA (Rao e Rao, 1995).

Redes de Multicamada *Feedforward*

As redes de multicamada *feedforward* podem possuir várias camadas de neurónios (Rao e Rao, 1995). Na Figura 2.8 é possível observar uma representação esquemática típica desta arquitetura de rede. A primeira camada é denominada como camada de entrada e a última camada é designada de camada de saída (Ding *et al.*, 2011; Kasabov, 1996; Lauwers *et al.*, 2013; Lingireddy e Brion, 2005). As setas a apontar para da camada de entrada representam os sinais de entrada ou os dados disponibilizados ao modelo neuronal. As setas que conectam os neurónios da primeira camada aos da camada oculta e as setas que ligam estes à camada de saída podem ser consideradas como o equivalente às sinapses biológicas. As setas a sair da camada de saída representam os sinais de saída ou a informação que as RNA concedem como resposta após processamento da informação (Lingireddy e Brion, 2005).

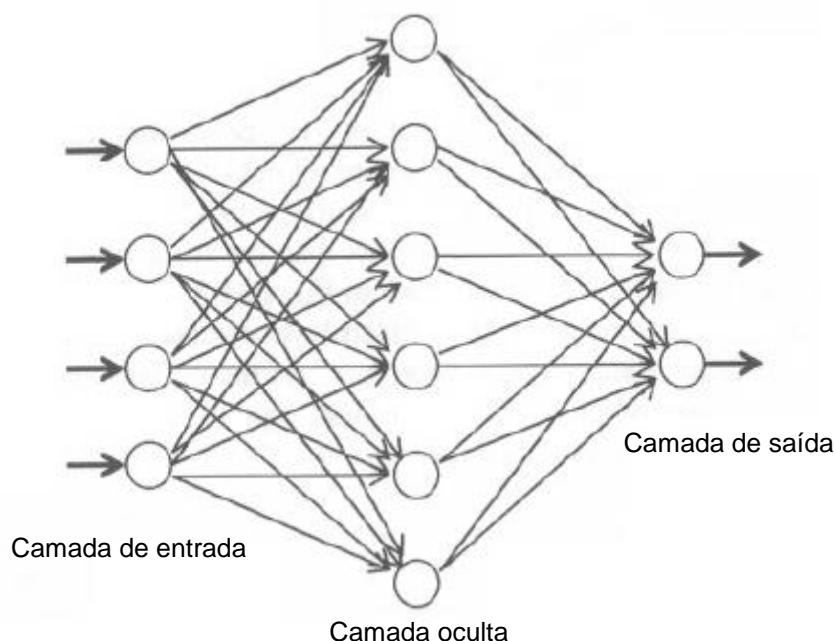


Figura 2.8 – Representação esquemática típica da arquitetura de uma rede de multicamada feedforward (adpatado de Lingireddy e Brion, 2005)

Apesar de na Figura 2.8 apenas estar representada uma camada oculta, estas redes neuronais podem conter uma ou mais camadas ocultas entre a camada de entrada e a camada de saída. Ao adicionar-se uma ou mais camadas ocultas, aumenta-se o número de camadas com neurónios de carácter computacional e é possível extrair estatísticas de ordem superior, o que é particularmente útil nos casos em que a quantidade de informação concedida à rede é bastante grande. Este modelo tal como no de camada única ao ser designado como *feedforward* significa que a transmissão dos estímulos de camada para camada ocorre apenas num sentido (Haykin, 1999).

Neste tipo de redes, o algoritmo de aprendizagem mais comumente usado é o algoritmo de retro propagação (*back-propagation*). Este algoritmo assume um determinado conjunto de pesos sinápticos no início e à medida que o treino ocorre, vai verificando o erro dos resultados face aos valores alvo. Se o erro estiver fora da tolerância permitida, o algoritmo altera os pesos sinápticos de modo a tentar corrigi-lo. Nas RNA *feedforward*, a determinação correta dos pesos sinápticos das conexões, é o fator central e mais importante na modelação (Lingireddy e Brion, 2005).

Redes Neuronais Recorrentes

Uma rede neuronal recorrente distingue-se de uma rede neuronal *feedforward*, se tiver pelo menos um *feedback loop*, ou seja, pelo menos um estímulo de saída de um neurónio é reenviado para trás, para servir como estímulo de entrada em neurónios de uma camada

anterior. (Haykin, 1999; Kasabov, 1996). Este modelo de RNA pode até possuir apenas uma camada de neurónios computacionais, onde cada um dos neurónios recambia sinais para outros neurónios da mesma camada, como é visível na Figura 2.9. Neste exemplo não existe nenhum estímulo recorrente de um neurónio para si mesmo, todavia esta situação também é viável de acontecer (Haykin, 1999).

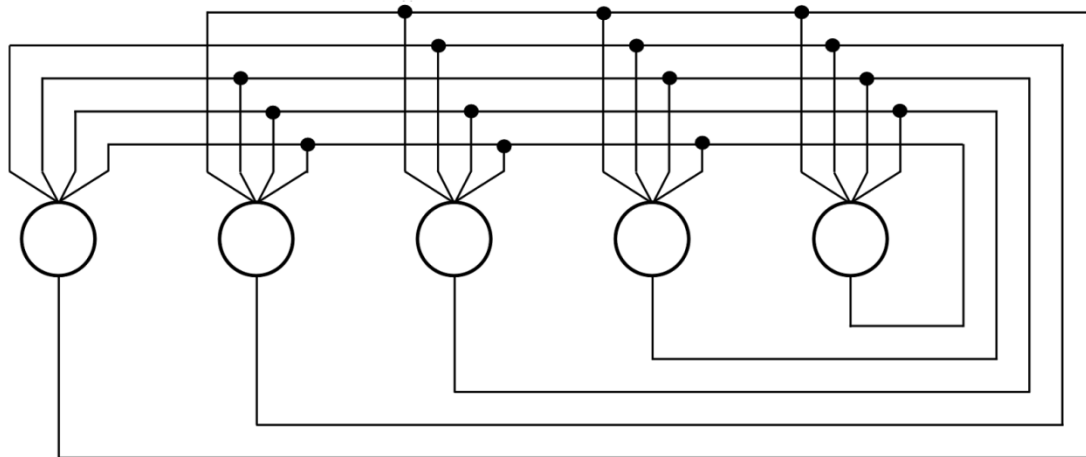


Figura 2.9 – Representação esquemática da arquitetura de uma rede recorrente de uma só camada (adpatado de Haykin, 1999)

Existem dois modelos principais de redes recorrentes, as redes recorrentes parciais e as redes totalmente recorrentes. Nos modelos recorrentes parciais, os estímulos que são reenviados para trás são seletivamente escolhidos, enquanto nas redes totalmente recorrentes, qualquer neurónio é livre de se conectar a outro (Kasabov, 1998). A vantagem das redes neuronais recorrentes sobre o modelo *feedforward* é o facto de se poder armazenar e reutilizar informação, durante um determinado período de tempo após esta ter sido produzida (Bose e Liang, 1996; Rojas, 1996). Esta característica das redes recorrentes permite-lhes resolver problemas que as redes *feedforward* não conseguem (Rojas, 1996).

Outros Tipos de Redes Neuronais

Existem outras arquiteturas de redes menos populares como as redes neuronais probabilísticas ou as redes de função de base radial. As redes neuronais probabilísticas, desenvolvidas para classificar padrões existentes na informação concedida à rede, são redes *feedforward* e possuidoras de um algoritmo de treino supervisionado, similar ao de retro propagação (Shan *et al.*, 2002). Este tipo de rede oferece as seguintes vantagens sobre as típicas *feedforward* (Specht, 1990):

- O treino é realizado de forma muito mais veloz;
- É possível adicionar a qualquer momento novos dados, o que permite continuar a desenvolver a rede durante a sua aprendizagem sem ser necessário reconverter toda a rede;

Na Figura 2.10, vislumbra-se a organização por camadas de uma rede neuronal probabilística, na execução da classificação de padrões em duas categorias. Os neurónios (designados como unidades) de entrada funcionam meramente como unidades de distribuição de informação. A informação é distribuída por todas as unidades padronizadas da primeira camada oculta, as quais são treinadas para reconhecerem os padrões e executarem uma operação não linear, antes de enviarem um sinal de ativação às unidades de soma. Cada unidade da segunda camada oculta soma os sinais de uma categoria de padrões treinados na primeira camada oculta, existindo neste caso duas categorias, também existem duas unidades de soma. A unidade de saída produz resultados binários, a partir de uma única variável calculada a partir do número de padrões treinados de cada uma das duas categorias (Specht, 1990).

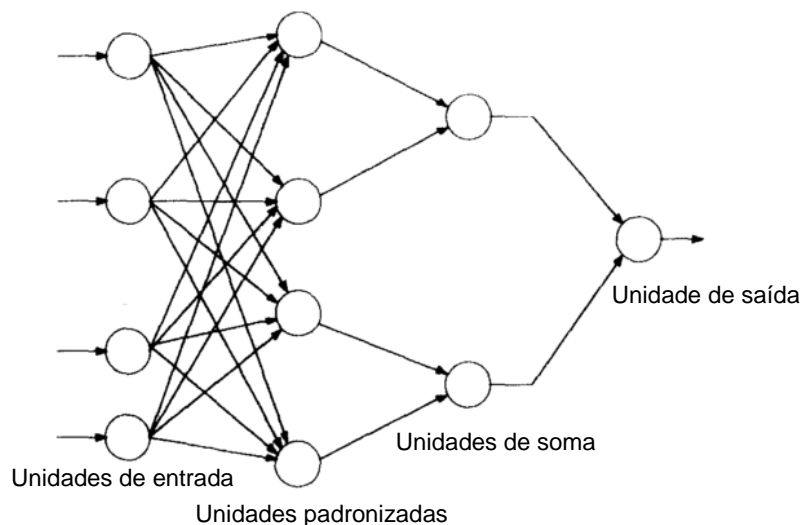


Figura 2.10 – Representação esquemática da arquitetura de rede probabilística (adaptado de Specht, 1990)

As redes de função de base radial podem ser consideradas como redes de duas camadas computacionais na sua forma mais simples, sendo usadas sobretudo no campo da investigação em análises numéricas (Haykin, 1999). Os neurónios da camada oculta contêm um parâmetro vetorial designado por centro, a partir do qual é medida uma distância até ao vetor de entrada. O resultado desta medição é processado por uma função não linear de uma única variável e enviado para os neurónios da camada de saída (Taylor, 1996). A camada de saída funciona de forma linear e submete a resposta final da rede de função de base radial. De acordo com esta metodologia, a aprendizagem é equivalente a encontrar o melhor espaço multidimensional para o vetor centro estar localizado, de forma a providenciar o melhor ajuste possível em termos de distância ao vetor de entrada. A generalização é equivalente à utilização desta superfície multidimensional para interpolar os dados de teste (Haykin, 1999). Na Figura 2.11 temos um exemplo da arquitetura típica deste tipo de RNA.

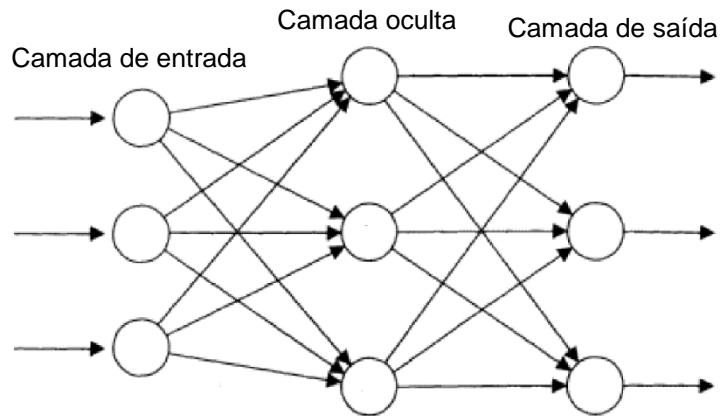


Figura 2.11 – Representação esquemática da arquitetura de rede de função de base radial (adaptado de Kasabov, 1996)

No cômputo global, apesar dos desenvolvimentos e bons resultados que se têm registado nesta área tecnológica, ainda há problemas por resolver. As teorias e topologias precisam de continuar a ser evoluídas e generalizadas, dado que muitas das investigações realizadas até ao momento, dizem respeito a casos especiais. Espera-se que nos próximos anos os algoritmos de aprendizagem e as respetivas arquiteturas de rede continuem a ser desenvolvidos, permitido um aumento na capacidade de aprendizagem das RNA e a construção de uma inteligência computacional cada vez mais apurada (Ding *et al.*, 2011).

3. Metodologia

3.1 Descrição Sumária

No presente trabalho, pretende-se através do recurso às redes neuronais, investigar o funcionamento e gestão de um digestor anaeróbio em dois casos de estudo diferentes, tendo em vista a otimização da produção de energia a partir do biogás produzido em cada digestor. O primeiro caso de estudo diz respeito ao digestor anaeróbio da ETAR do Seixal, enquanto no segundo caso de estudo, o digestor está situado na ETAR de Sesimbra. As duas ETAR em questão pertencem ao Grupo AdP. Em ambos os casos, é tida em consideração a metodologia utilizada na investigação anteriormente realizada na ETAR do Seixal no mesmo âmbito do estudo em questão no ano de 2014 (Pinto, 2014).

3.2 Entidades Envolvidas nos Casos de Estudo

3.2.1 Grupo Águas de Portugal

O Grupo AdP integra o setor empresarial do Estado e é um dos grandes grupos do setor ambiental em Portugal. O grupo tem um papel importantíssimo no domínio do abastecimento público de água e do saneamento de águas residuais, ao servir 95% da população com sistema público de abastecimento de água e 90% da população com sistema público de saneamento de águas residuais urbanas. O tratamento e valorização de resíduos é também uma das áreas em que o grupo tem maior raio de ação (AdP, 2015).

A AdP foi constituída em 1993, tendo como principal objetivo desenvolver sistemas multimunicipais de abastecimento de água e de saneamento de águas residuais. No ano de 2000, participações do Estado português passaram a integrar o grupo AdP, nomeadamente a Empresa Geral do Fomento (EGF), dando possibilidade ao grupo de entrar no ramo de tratamento e valorização de resíduos sólidos urbanos. Recentemente, o modelo institucional inicial foi reformulado, dando origem ao atual modelo, assente em contratos de parceria entre o Estado e as autarquias locais. (AdP, 2015).

No presente, o Grupo AdP agrega mais de 40 empresas, que para além de atuarem nas áreas de ação já referidas, atuam também na área das energias renováveis, serviços partilhados e em mercados internacionais. A nível nacional, o grupo AdP, em associação com os municípios, serve mais de 80% da população residente em Portugal, com principal incidência na esfera de abastecimento de água, saneamento de águas residuais e de tratamento e valorização de resíduos sólidos urbanos (AdP, 2015). Na figura 3.1, é possível observar a localização das empresas de abastecimento de água e saneamento de águas residuais no início do ano de 2015, antes de ocorrer uma reestruturação institucional no Grupo AdP.

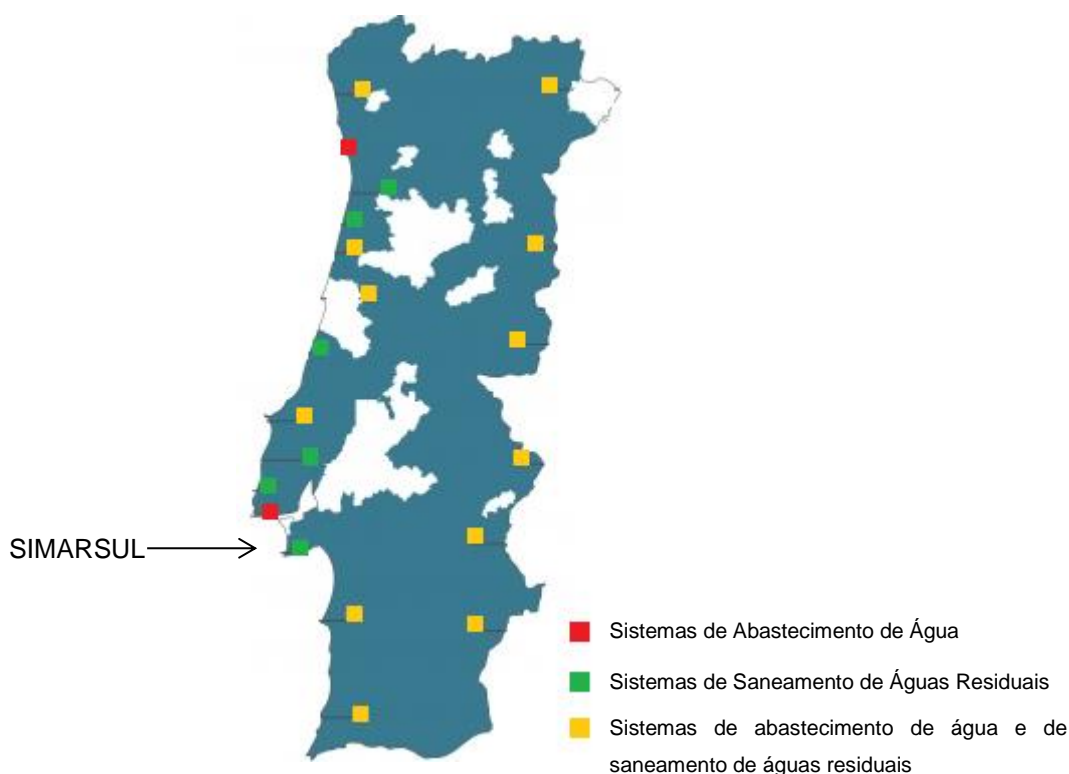


Figura 3.1 – Empresas de abastecimento de água e saneamento de águas residuais (adaptado de AdP, 2015)

Além do elevado número de empresas que constituem o Grupo AdP, a grande dimensão do Grupo pode constatar-se também através da observação da Tabela 3.1, onde estão dispostos os dados de atividade da empresa relativos ao ano de 2014. A nível de abastecimento e tratamento de água, os números registados advêm do funcionamento de 79 estações de tratamento de água (ETA), 967 ETAR, mais de 2 200 Estações Elevatórias e uma extensão de redes de abastecimento e de saneamento de cerca de 20 mil quilómetros (SIMRIA, 2015).

Tabela 3.1 – Dados de atividade do Grupo Águas de Portugal em 2014 (adaptado de AdP, 2015)

Dados de Atividade	Ano (2014)
Volume de água produzida (milhões de m ³)	573,3
Volume de água distribuída (milhões de m ³)	74,4
Volume de águas residuais tratadas (milhões de m ³)	546,7
Volume de resíduos sólidos urbanos tratados (milhões de toneladas)	3,1
Número de colaboradores	4788

Desde o ano de 2009, a estratégia e os planos de investimento do Grupo visam uma aposta em energias renováveis e na maximização do aproveitamento de ativos e recursos endógenos, como os subprodutos de tratamento de águas. A melhoria no aproveitamento de biogás, que advém do processo de digestão anaeróbia, é uma das áreas que tem sido alvo de um reforço

de investimento num passado recente (AdP, 2015). Exemplo do interesse da AdP nesta área é o projeto iniciado no ano de 2013, em associação com o Instituto Superior de Engenharia do Porto, para aumentar a produção de energia num digestor de uma ETAR da SIMRIA recorrendo ao uso de RNA. O estudo revelou o potencial das RNA para cumprir o objetivo proposto, levando a que no ano de 2014 o projeto se abrisse a mais entidades. Uma das empresas do Grupo AdP escolhidas para aderir ao projeto foi a SIMARSUL, que, em colaboração com a FCT, iniciou a aplicação deste projeto na ETAR do Seixal (Pinto, 2014).

3.2.2 SIMARSUL

Antes da Reestruturação

A SIMARSUL é uma sociedade anónima de capitais públicos formada em 2003 e faz parte do conjunto de empresas que constituem o Grupo AdP. Esta empresa possui a concessão exclusiva por 30 anos (com início em Dezembro de 2004) da atividade de recolha, tratamento e rejeição de águas residuais em oito municípios da Península de Setúbal, os quais podem ser observados na Figura 3.2. A área geográfica total servida é de 1450 km². O principal objetivo da SIMARSUL passa por contribuir para a preservação e melhoria da qualidade do ambiente e dos recursos hídricos, de modo a elevar a qualidade de vida dos habitantes da Península de Setúbal (SIMARSUL, 2015).



Figura 3.2 – Municípios abrangidos pela SIMARSUL (SIMARSUL, 2015)

O Sistema Multimunicipal de Saneamento gerido pela SIMARSUL, que atingirá a sua capacidade máxima de funcionamento no ano de 2034, está dimensionado para assegurar o tratamento de um caudal médio diário de 160 mil m³ de águas residuais, correspondente a uma população de 1,4 milhões de habitantes-equivalentes. A solução técnica do Sistema

Multimunicipal está assente em subsistemas de drenagem e saneamento, sendo que a cada um corresponde uma ETAR, várias estações elevatórias e um sistema de drenagem gravítico e elevatório que engloba emissários, interceptores e condutas elevatórias (SIMARSUL, 2015). Os números relativos ao sistema em exploração e ao sistema a funcionar na sua capacidade máxima podem ser observados nas Tabelas 3.2 e 3.3.

Tabela 3.2 – Dados do sistema em exploração (adaptado de SIMARSUL, 2015)

Descrição	Em exploração	População servida em 2013 (habitantes equivalentes)
Número de ETAR	21	
Número de Estações Elevatórias	96	446 000
Extensão de Emissários, interceptores e condutas elevatórias (km)	344	

Tabela 3.3 – Dados do sistema relativos ao ano horizonte de projeto (adaptado de SIMARSUL, 2015)

Descrição	Intervenção			Total	População servida em 2034 (habitantes equivalentes)
	Integradas	Construídas	Remodeladas		
Número de ETAR	0	16	12	28	
Número de Estações Elevatórias	15	83	32	130	
Extensão de Emissários, interceptores e condutas elevatórias (km)	155	250	15	420	1 400 000

Reestruturação da Empresa

Durante os últimos meses da realização da presente dissertação, o Grupo AdP procedeu a uma reestruturação do seu modelo institucional. A SIMARSUL foi agregada a outras empresas do setor e pertencentes ao Grupo AdP, de modo a formar a empresa Águas de Lisboa e Vale do Tejo, S.A. A empresa SIMARSUL - Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal, SA foi extinta, transferindo-se, ao abrigo do disposto no Decreto-lei n.º 94/2015, de 29 de maio, todos os direitos e obrigações dessa Sociedade Concessionária para a nova empresa, sendo que a gestão do Sistema está delegada na Empresa Portuguesa das Águas Livres, SA.

3.3 Casos de Estudo

No ano de 2014, como já mencionado, a ETAR do Seixal foi um dos locais escolhidos para a continuação da investigação do projeto desenvolvido pela AdP, que tem por âmbito estudar a produção de biogás em ETAR recorrendo ao uso de RNA como ferramenta de apoio. Nessa

altura chegou-se a equacionar a entrada da ETAR de Sesimbra no estudo, o que não veio a concretizar-se devido ao facto de os dados disponíveis não estarem em formato digital e serem em menor número, em comparação com a ETAR do Seixal. Volvido um ano, o número de dados disponíveis aumentou, pelo que o presente trabalho além de continuar a pesquisa na ETAR do Seixal, inicia também a investigação na ETAR de Sesimbra.

3.3.1 Caso de Estudo - ETAR do Seixal

A construção da ETAR do Seixal iniciou-se no ano de 2007, na freguesia de Aldeia de Paio Pires, concelho do Seixal, num local anteriormente ocupado por escórias da Siderurgia Nacional. Na Figura 3.3, é possível observar a localização da ETAR, no interior da freguesia de Aldeia de Paio Pires (limites contornados a preto) no concelho do Seixal. A ETAR foi dimensionada para servir grande parte da população do concelho do Seixal, num total de 110 mil habitantes em ano de arranque e um total de 156 mil habitantes em ano horizonte de projeto. Na realização da empreitada, foi investido um total de 10 291 248,50 de euros, dos quais 3 215 445,74 de euros advêm de um financiamento do QREN, do Programa Operacional Temático Valorização do Território (POVT) e da UE (SIMARSUL, 2015).



Figura 3.3 – Concelho do Seixal e local de Implantação e vista aérea da ETAR

O subsistema de drenagem e saneamento do Seixal, além da ETAR, possui sete estações elevatórias e 31 km de emissários e condutas elevatórias (SIMARSUL, 2015). Os dados de base (caudais e cargas) considerados para efeitos de dimensionamento podem ser consultados na Tabela 3.4.

Tabela 3.4 – Dados de base da ETAR do Seixal (adaptado de FERROVIAL-AGROMAN/CADAGUA, 2007)

Parâmetro	Unidades	Valor considerado	
		Ano de Arranque	Ano Horizonte
Caudal médio diário de época baixa	m ³ /dia	25 100	42 050
Caudal médio diário de época alta	m ³ /dia	26 440	43 850
Caudal de ponta de época baixa	m ³ /hora	2 200	3 600
Caudal de ponta de época alta	m ³ /hora	1 620	2 660
CBO ₅	Época baixa	5 600	7 940
	Época alta	6 810	9 340
CQO	Época baixa	11 200	15 880
	Época alta	13 620	18 680
Sólidos Suspensos Totais (SST)	Época baixa	8 410	11 900
	Época alta	10 220	14 000

Sistema de tratamento:

- Linha de tratamento da fase líquida (limites da área afeta contornados a vermelho na Figura 3.4);
- Linha de tratamento da fase sólida (limites da área afeta contornados a azul na Figura 3.4);
- Serviços auxiliares.

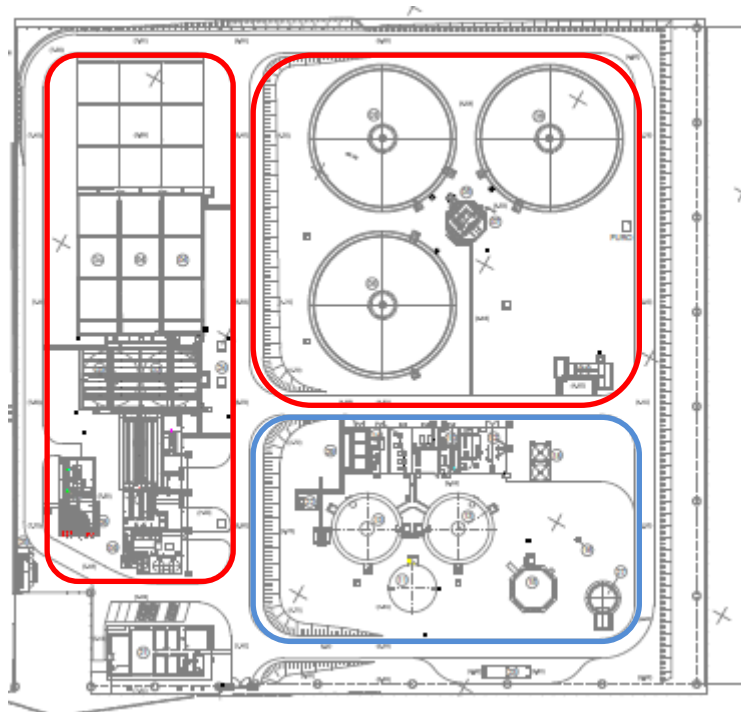


Figura 3.4 – Planta do sistema de tratamento da ETAR do Seixal (adaptado de FERROVIAL-AGROMAN/CADAGUA, 2007)

Linha de tratamento da fase líquida (FERROVIAL-AGROMAN/CADAGUA, 2007):

- Caudal afluyente à obra de entrada
 - Cerca de 20% do caudal que chega à obra de entrada aflui por um coletor gravítico, sendo posteriormente elevado dentro da instalação, enquanto o restante caudal advém por bombagem;
 - Obra de entrada composta inicialmente, por uma câmara de receção e dissipação de energia que recebe a água residual bombeada. À saída da câmara, ocorre a mistura entre o afluyente bombeado e o gravítico que foi elevado.
- Pré-Tratamento
 - A gradagem/tamisação é efetuada por dois tamisadores com espaçamento de malha de 3 mm e de limpeza automática, colocados em paralelo em canais de secção retangular. Existe ainda um canal de *bypass* com uma grelha de limpeza manual e de 10 mm de espaçamento;
 - Os gradados são transportados e compactados por um parafuso/compactador, que funciona em sincronia com os tamisadores, para um contentor “multibenne” de 5m³ de capacidade;
 - A desarenação/desengorduramento é realizada em duas linhas em paralelo. A extração de areia dos desarenadores é efetuada por duas bombas (uma por desarenador), sendo de seguida enviada para um classificador de areias. As gorduras provenientes dos desarenadores são recolhidas e concentradas pelo sistema de desengorduramento, juntamente com as gorduras dos decantadores primários. No fim da etapa da etapa de concentração de gorduras, estas são elevadas para a câmara de homogeneização de lamas (linha de tratamento da fase sólida).
- Tratamento Primário
 - O tratamento primário é executado por dois decantadores lamelares com mecanismo de raspagem e espessamento de lamas de acionamento central. Os decantadores (15 m x 15 m) são de geometria quadrada, de soleira inclinada e circular. As lamelas estão dispostas em módulos em forma de colmeia, totalmente submersos, perfazendo um ângulo de inclinação de 60°. Os sobrenadantes dos decantadores primários são extraídos e enviados para um concentrador de gorduras. A extração das lamas primárias espessadas ocorre de forma temporizada com recurso a bombas de parafuso, que as enviam para uma câmara de homogeneização.
- Tratamento Secundário
 - O tratamento secundário divide-se em três linhas de tratamento de igual capacidade. Cada linha possui um reator biológico e um decantador secundário;
 - Os reatores são de lamas ativadas com arejamento por ar difuso, funcionando em regime de média carga, de forma contínua e com uma configuração de tipo fluxo

pistão. O volume total dos reatores biológicos é de 7000 m³. Cada reator incorpora um seletor biológico anóxico à cabeça de modo a controlar a biomassa filamentosa. Existe a possibilidade de incorporar futuramente câmaras anóxicas e arejadas para remoção de azoto e fósforo, caso necessário;

- Os decantadores secundários são de fundo plano, de planta circular e detêm um diâmetro de 36 m. Estão equipados com uma ponte raspadora, a qual tem um raspador de fundo que arrasta as lamas até tubos de sucção e um raspador de superfície para remoção de escumas. As lamas são conduzidas a uma estação elevatória, a partir da qual são enviadas por bombagem para o reator biológico (recirculação externa) ou para um espessador mecânico (lamas em excesso). As escumas formadas á superfície são retidas pelo defletor periférico e depois recolhidas para uma tremonha, a partir da qual seguem por gravidade até um poço de bombagem, sendo posteriormente enviadas para espessamento mecânico.

➤ Tratamento Terciário

- Este tratamento consiste num sistema de desinfecção por radiação ultravioleta (UV) de limpeza automática ajustável. O sistema inclui a incorporação de balastros eletrónicos nos módulos de lâmpadas de baixa pressão e de eixo paralelo ao sentido de escoamento, as quais podem funcionar a diferentes níveis de potência;
- A água tratada é depois encaminhada até um emissário de descarga que desagua no estuário no Tejo.

Linha de tratamento da fase sólida (FERROVIAL-AGROMAN/CADAGUA, 2007):

➤ Espessamento de lamas

- As lamas primárias são espessadas nos decantadores primários. A purga de lamas ocorre, de forma temporizada, e pode ser regulada por um operador, consoante o grau de espessamento desejado, tendo em conta a manutenção das eficiências de remoção do órgão. Após a purga, as lamas são enviadas por bombagem para a câmara de homogeneização;
- As lamas em excesso produzidas durante o tratamento secundário são extraídas diretamente do poço de recirculação de lamas, através de uma estação elevatória. A escolha para o espessamento destas lamas recaiu sobre um tambor de espessamento de estrutura tubular cilíndrica. O tambor está equipado com uma tela filtrante e é acionado por moto-redutor elétrico de baixo consumo. Durante o espessamento, é adicionado polieletrólito, de modo a favorecer a formação de flocos. As lamas espessadas escoam graviticamente para o tanque de homogeneização;
- Na câmara de homogeneização, com um volume de 45 m³, as lamas primárias e as lamas em excesso são misturadas por um agitador. O tempo de retenção, caso a

câmara esteja a funcionar a caudal máximo, pode ultrapassar as quatro horas. As lamas homogeneizadas são elevadas da câmara para os digestores anaeróbios.

➤ Estabilização anaeróbia

- A estabilização anaeróbia ocorre através de um processo de digestão anaeróbia de alta carga em gama de temperaturas mesófilas. Na ETAR, operam dois digestores de 18 m de diâmetro, com um volume unitário de 2149 m³. A agitação e o aquecimento de lamas efetuam-se através da tecnologia HEATMIX. Este sistema é caracterizado pela colocação de unidades de agitação e aquecimento no exterior do digestor. As unidades dispostas permitem agitar a massa de lamas através de um efeito *air-lift*, gerado por injeção de gás em pressão num tubo interior, enquanto na camisa exterior ao tubo circula água quente. A água é aquecida a partir do funcionamento de uma caldeira, possuidora de um queimador misto de biogás e gás natural, ou com recurso à água resultante do sistema de refrigeração do cogenerator;
- Para a promoção da recirculação de lamas, são utilizados três permutadores-agitadores em cada digestor, que forçam uma ação giratória através de um compressor, mantendo assim as partículas em suspensão;
- As lamas digeridas são armazenadas num depósito tampão com volume útil de 763 m³ e um diâmetro de 12 m. O depósito possui um agitador que homogeneiza as lamas;
- Existe também uma linha de biogás, para que este seja usado como fonte de energia. Os equipamentos principais que compõem a parte de armazenamento e rede de gás em baixa pressão são:
 - Um gasómetro de dupla membrana de 13,1 m de diâmetro e de 1000 m³ de volume, com uma capacidade de armazenamento sobre a produção diária no ano horizonte de gás de 6 h/dia;
 - Uma tocha de queima com uma capacidade de 459 Nm³/h.

Os equipamentos principais que constituem a parte referente à fase de recuperação de energia são:

- Motogerador GUASCOR que produz energia elétrica e aproveita a energia térmica para aquecimento dos digestores;
- Dois permutadores e correspondentes bombas de água de refrigeração, que aproveitam o calor residual da água das camisas;
- Sistema de eliminação de sulfídrico (H₂S) no biogás de digestão através da injeção de cloreto férrico na câmara de mistura de lamas, de forma a evitar a corrosão do motogerador.

➤ Desidratação de lamas

- O sistema de desidratação é composto por duas centrífugas que devem operar em média sete horas por dia, durante cinco dias por semana, para as condições previstas no ano de arranque. No ano horizonte de projeto, caso se confirmem as

cargas previstas, as centrífugas irão trabalhar durante dez horas por dia, durante cinco dias por semana, ou será adquirida uma terceira centrífuga que permitirá reduzir as horas de funcionamento de cada uma;

- A alimentação das centrífugas é feita por dois grupos elevatórios de parafuso excêntrico;
- Após a adição de polieletrólito à lama a desidratar, esta é introduzida no rotor e, devido ao efeito da força centrífuga, separa-se em duas fases (sólido e líquido). Os sólidos são extraídos pela ação contínua de um parafuso transportador sem-fim e conduzidos a um grupo elevatório, que os envia para um dos dois silos existentes de 120 m³.

Serviços auxiliares (FERROVIAL-AGROMAN/CADAGUA, 2007):

➤ Linha de Tratamento de odores

- O sistema de tratamento de odores funciona com base nos seguintes critérios:
 - Eliminação de odores por processo químico, o qual inclui torres para lavagem que possuem uma solução ácida, uma oxidante e uma básica;
 - A extração do ar realiza-se por condutas de ventilação;
 - Renovações de ar nos espaços confinados visitáveis e nos espaços confinados não visitáveis.
- As instalações que contêm este sistema de tratamento são o edifício de pré-tratamento, o depósito tampão de lamas espessadas, o edifício de desidratação e os silos de lamas desidratadas.

➤ Reservatório de água tratada

- O reservatório está colocado a jusante da desinfecção e serve para armazenar a água, que será posteriormente reutilizada no suprimento das necessidades ao funcionamento da ETAR. Existe uma etapa de filtração antes do armazenamento da água no reservatório.

➤ Central Hidropneumática

- Central hidropneumática com dois grupos elevatórios que levam a água às restantes instalações de modo a suprir as necessidades da ETAR.

➤ Laboratório

- Laboratório onde poderão ser realizadas as análises comuns e necessárias para um bom funcionamento da ETAR.

Desenvolvimento do estudo

A metodologia empregue neste primeiro caso de estudo é apresentada na Figura 3.5.

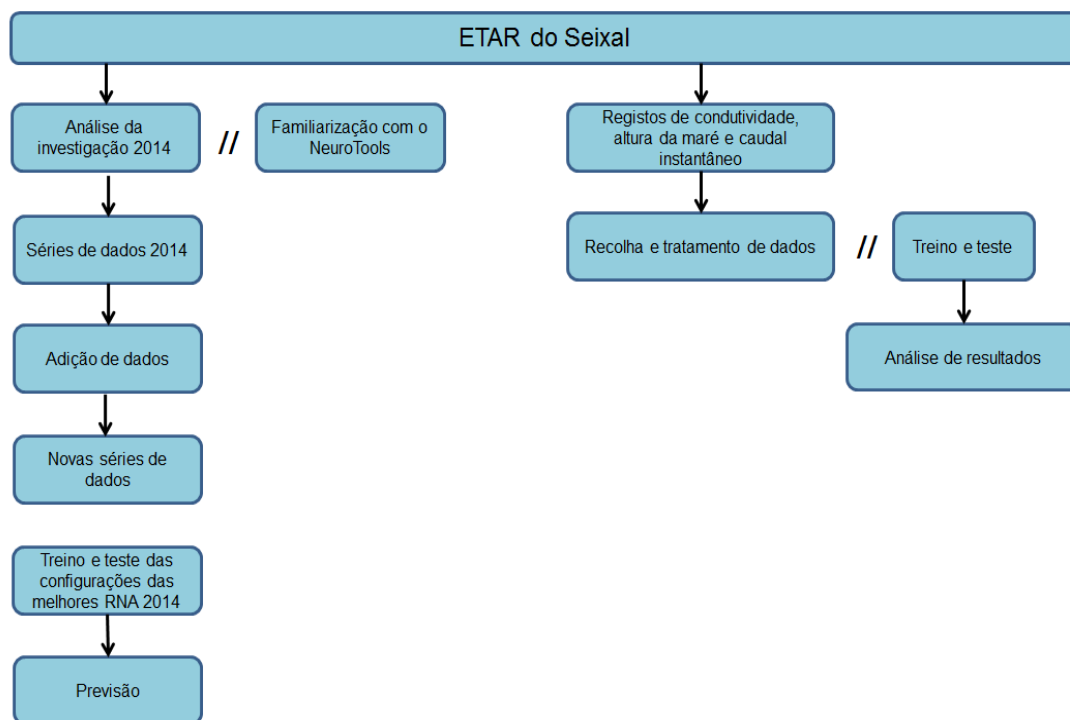


Figura 3.5 – Metodologia aplicada na investigação realizada na ETAR do Seixal

O primeiro caso de estudo diz respeito à segunda fase de desenvolvimento da investigação iniciada no ano de 2014 na ETAR do Seixal, sendo que neste ano acresce a componente condutividade como um dos fatores principais a ter em conta na modelação do funcionamento do digestor anaeróbio. O passo inicial a tomar foi executar uma análise cuidada da investigação já realizada, na qual se teve em especial atenção alguns aspetos, de seguida enunciados:

- Um dos aspetos a que se deu maior importância foi a metodologia utilizada, de modo a que o atual trabalho fosse produzido em consonância com o anterior. A conformidade entre os dois trabalhos era fulcral, pois só assim seria possível aproveitar tudo o que de bom tinha sido feito e gerar melhores resultados;
- Em segundo lugar, foi importante fazer uma análise dos resultados alcançados e respetivas conclusões do estudo. Esta análise serviu para se ter uma perceção mais rápida do potencial das RNA no presente caso de estudo;
- Por último, teve-se em consideração tudo aquilo que correu bem e menos bem na investigação anterior. A constatação destes pormenores permite evitar erros desnecessários e apostar mais no que demonstrou um maior potencial de evolução.

Tendo em conta que o programa de computador a utilizar para produção das redes neuronais se encontrava à partida selecionado, e os dados tratados da investigação anterior disponibilizados, iniciou-se numa fase prévia a etapa de familiarização com o NeuralTools®. Durante este período, procurou-se conhecer as funcionalidades da ferramenta a utilizar e aumentar a capacidade de manuseamento da mesma. Na conclusão deste processo, recriaram-se algumas das melhores redes neuronais geradas no ano de 2014, comparando-se os resultados, de forma a ser possível comprovar se a utilização do programa no momento estava a ser correta.

A terceira etapa passou pela recolha, preparação e adição de dados às séries de variáveis já existentes e transpostas do primeiro trabalho para este. Os novos dados disponibilizados encontravam-se registados em formato digital, o que simplificou a sua recolha. Apesar de a tarefa a nível da escolha das variáveis a usar nos testes estar facilitada pela investigação anterior, foi efetuado um estudo e uma análise das variáveis disponíveis. As condições em que as redes neuronais foram concebidas na primeira fase de desenvolvimento do projeto na ETAR do Seixal foram estudadas aquando da adição de novos valores. Geraram-se gráficos de cada variável, de modo a verificar a existência de oscilações nas séries e possíveis valores fora da matriz normal de resultados, que obrigassem à sua remoção. A verificação das correlações entre variáveis foi realizada através do cálculo do coeficiente de correlação de Pearson (r), dado ser necessário tomar conhecimento do impacto provocado pelos novos valores adicionados nas séries de dados. O cálculo do coeficiente de Pearson permite conhecer a força e a direção (decrecente ou crescente, caso o resultado seja negativo ou positivo, respetivamente) da relação linear entre duas variáveis X e Y (Ahlgren e Jarneving, 2003). O coeficiente pode ser calculado a partir das seguintes expressões matemáticas (Puth *et al*, 2014):

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (\text{equação 3.1})$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad (\text{equação 3.2})$$

$$r = \frac{\sum_{i=1}^N \{(X_i - \bar{X})(Y_i - \bar{Y})\}}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (\text{equação 3.3})$$

Finalizada a preparação das séries de dados, após a adição de valores, executaram-se novos ensaios, com o objetivo de conceber RNA capazes de prognosticar valores diários de biogás produzido na etapa de digestão anaeróbia. Nos testes executados, foi dada total primazia ao conjunto de variáveis que demonstrou um maior potencial na investigação efetuada

anteriormente. Esta opção teve por base o facto de não haver tempo de repetir o vasto conjunto de ensaios realizado no ano anterior, visto esse não ser o objetivo principal do atual trabalho. A seguinte fase foi a realização da verificação e análise dos resultados produzidos a partir das novas séries de dados. Esta análise era de extrema importância, dado o principal impeditivo para a génese de melhores resultados no trabalho realizado em 2014 ser o facto de o conjunto de dados disponíveis ser curto, aspeto que foi amenizado após o acréscimo de mais valores. Esta etapa marcou o culminar da fase mais simples do atual trabalho, pois até aqui procurou-se apenas aproveitar e potenciar aquilo que de muito bom se tinha concebido no passado.

A influência da salinidade no funcionamento do digestor era o principal foco de atenção na ETAR do Seixal no presente ano, pelo que a recolha e preparação dos dados de condutividade assinalava o início desta nova fase do estudo. A recolha de dados de altura de maré também foi de extrema importância, visto que se trata de uma variável com grande influência nos valores de condutividade das águas residuais afluentes à ETAR. A recolha de dados não foi muito difícil, dado que estes se encontravam na sua totalidade em formato digital. O impacto da altura de maré na condutividade registada levou à necessidade da execução de curvas de maré. Ao assumir-se que a onda de maré é sinusoidal, o que não é rigorosamente verdade, é possível determinar a altura da maré a qualquer instante recorrendo às equações 3.4 e 3.5 (Instituto Hidrográfico, 2015):

A altura de água (Y) em qualquer momento depois de uma preia-mar (PM) pode ser calculada com base na seguinte equação:

$$Y = \frac{H+h}{2} + \frac{H-h}{2} \cos \frac{\pi \cdot t}{T} \quad (\text{equação 3.4})$$

A altura de água (Y1) em qualquer momento depois de uma baixa-mar (BM) pode ser calculada com base na seguinte equação:

$$Y1 = \frac{h+H1}{2} + \frac{h-H1}{2} \cos \frac{\pi \cdot t1}{T1} \quad (\text{equação 3.5})$$

Onde:

- Os valores de H e H1 representam as alturas de água da PM e BM, respetivamente, as quais enquadram o intervalo de tempo no qual se vai efetuar o cálculo;
- Os valores de T e T1 representam o intervalo em tempo entre uma PM e BM e entre uma BM e PM, respetivamente;
- Os valores de t e t1 representam o intervalo em tempo entre uma PM e a hora a que se pretende saber a altura de maré e o intervalo em tempo entre uma BM e a hora a que se pretende saber a altura de maré, respetivamente.

De seguida, realizou-se o tratamento de dados disponíveis do conjunto de variáveis afetas à condutividade e prepararam-se as séries de dados. Concluída a etapa anterior, efetuou-se o cálculo do coeficiente de correlação de Pearson, de forma a se ter uma melhor noção da qualidade de valores disponíveis e da correlação entre variáveis.

O processo de preparação das séries de dados e respetiva análise de correlações afeto à condutividade ocorreu algumas vezes. Foi necessário repetir o procedimento, dado que os registos, de onde se retiraram os valores, foram disponibilizados de forma intermitente pela SIMARSUL, visto ainda estarem a ser trabalhados no decurso deste trabalho. À medida que os valores chegavam e eram acomodados nas séries de dados, geravam-se em seguida RNA com o intuito de prever valores de condutividade. A evolução dos resultados foi avaliada, o que conduziu várias vezes a que ocorresse uma seleção de dados, ou seja, só determinados valores sob certas condições incorporavam as séries que alimentavam a produção das RNA. Terminada a fase de criação das várias RNA, realizou-se uma análise final dos resultados produzidos. O último passo tomado na investigação referente a este caso de estudo foi a execução de um balanço da concretização dos objetivos iniciais propostos.

3.3.2 Caso de Estudo - ETAR de Sesimbra

A ETAR de Sesimbra foi construída no ano de 2000 na freguesia do Castelo, concelho de Sesimbra, e entrou em funcionamento no ano de 2001. Na Figura 3.6, é possível observar a localização da ETAR no interior da freguesia do Castelo (limites contornados a preto), no concelho de Sesimbra. No ano de 2007, a ETAR foi alvo de algumas remodelações na ordem dos 1 157 604,74 de euros. A infraestrutura está preparada para assegurar o tratamento de águas residuais urbanas, geradas por 18 mil habitantes no ano de arranque e 30 mil habitantes no ano horizonte de projeto. O Subsistema de Sesimbra, além da ETAR, contempla cerca de 11 km de emissários, duas estações elevatórias e condutas elevatórias (SIMARSUL, 2015).



Figura 3.6 – Concelho de Sesimbra e local de implementação e vista aérea da ETAR

Os dados de base (caudais e cargas) considerados no dimensionamento da instalação são apresentados na Tabela 3.5.

Tabela 3.5 – Dados de base da ETAR de Sesimbra (adaptado de SIMARSUL 2007)

Parâmetro	Unidades	Valor considerado	
		Ano de Arranque	Ano Horizonte
Caudal médio diário afluente	m ³ /dia	2720	6000
Caudal de ponta afluente	m ³ /hora	432	864
CBO ₅	kg/dia	971	1620
CQO	kg/dia	1942	3240
SST	kg/dia	1621	2700

Sistema de tratamento (SIMARSUL, 2007):

- Linha de tratamento da fase líquida (limites da área afeta contornados a vermelho na Figura 3.7);
- Linha de tratamento da fase sólida (limites da área afeta contornados a azul na Figura 3.7);
- Serviços auxiliares.

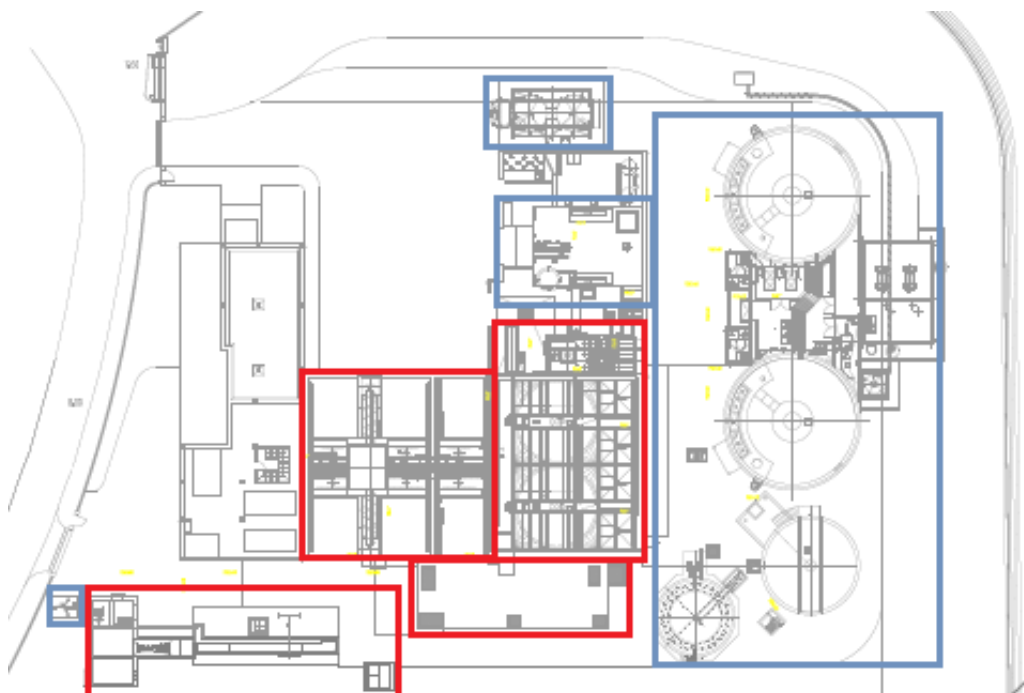


Figura 3.7 – Planta do sistema de tratamento da ETAR de Sesimbra (adaptado de Simarsul, 2007)

Linha de tratamento da fase líquida (SIMARSUL, 2007):

- Caudal afluyente à obra de entrada
 - O caudal de águas residuais chega por bombagem à obra de entrada, onde passa por um medidor de caudal e entra numa caixa aberta, sendo distribuído em seguida pelos dois canais de gradagem de forma equitativa;
 - Na zona central da caixa aberta recetora de caudal, encontra-se o sistema de *bypass* geral, que é controlado por uma caixa descarregadora de nível.
- Pré-Tratamento
 - O afluyente bruto passa por uma etapa de gradagem/tamisação, a qual é realizada por dois tamisadores do tipo “*aquaguard*” de 3 mm de espaçamento de malha. No canal de *bypass*, existe um tamisador mecânico do tipo “*step screen*” com o mesmo intervalo de malha dos instalados nos canais principais. Em ambos os casos, o equipamento é comandado por uma bóia de nível instalada a montante da gradagem, a qual é acionada pelo nível da linha líquida;
 - Os resíduos tamisados são conduzidos para um compactador de resíduos e posteriormente enviados para um contentor de 0,7 m³ de capacidade;
 - A ETAR possui dois órgãos designados de SEDIPAC 3D que podem funcionar em paralelo ou individualmente, consoante o caudal afluyente. Este órgão realiza duas operações de pré-tratamento (desarenação e desengorduramento) e o tratamento primário;
 - A partir de uma injeção de ar comprimido por difusores de bolha média, dá-se a separação das areias da matéria orgânica envolvente, as quais entram numa tubagem que as conduz até um classificador de areias;
 - No interior de duas cubas, desenrola-se a etapa de desengorduramento. As gorduras são encaminhadas para um poço e posteriormente enviadas para um espessador de gorduras, antes de serem despejadas num contentor.
- Tratamento primário
 - Os decantadores primários, englobados no SEDIPAC 3D, são decantadores lamelares com 7 m de diâmetro e uma área lamelar de 31 m². Os decantadores possuem uma ponte raspadora de fundo que arrasta as lamas até uma fossa central, onde estas são bombadas para um espessador. Após a decantação, o efluente é recolhido e conduzido até uma caleira central, onde à saída se encontra um tamisador tipo “*Aquaguard*” com 1mm de espaçamento. Depois de passar pelo tamisador, o efluente, por ação da gravidade, transita para o tratamento secundário.
- Tratamento secundário
 - O tratamento biológico é realizado com recurso à biofiltração, efetuada por duas baterias de BIOFOR, o qual possui uma área unitária de 22,7 m²;

- O BIOFOR é um reator biológico de leito fixo emerso com fluxos de ar e de água em co-corrente, em que as bactérias estão fixas sobre uma única camada de material granular. A primeira bateria de BIOFOR aplica um desbaste no carbono, enquanto a segunda bateria completa a tarefa a nível da remoção de carbono e realiza uma nitrificação parcial. Caso haja o perigo de o caudal disponível não garantir uma velocidade mínima de atravessamento do BIOFOR, existe uma bomba com a capacidade para fazer o caudal recircular, enviando-o para a entrada do órgão ou para a obra de entrada. As águas de lavagem do BIOFOR são recolhidas para o interior de uma cisterna, onde duas bombas submersíveis enviam as águas para o canal de entrada dos SEDIPAC 3D. Após a etapa de biofiltração, o efluente segue para o tratamento terciário, enquanto as lamas produzidas são remetidas por bombagem para os SEDIPAC 3D onde se misturam com as lamas primárias.
- Tratamento terciário
- O tratamento terciário é executado através de uma desinfecção por UV. Os seis módulos de 40 lâmpadas existentes estão instalados verticalmente em canal aberto e são de limpeza automática;
 - O efluente tratado é descarregado no Oceano Atlântico através de um emissário submarino.

Linha de tratamento da fase sólida (SIMARSUL, 2007):

- Espessamento de lamas
- As lamas provenientes do tratamento primário e do tratamento secundário, após se misturarem no SEDIPAC 3D, são extraídas para um espessador gravítico de 8 m de diâmetro e 50 m² de área, por duas bombas de parafuso excêntrico;
 - As lamas espessadas são retiradas do fundo do espessador gravítico por duas bombas de parafuso excêntrico que as enviam para a etapa de estabilização anaeróbia.
- Estabilização anaeróbia
- Esta etapa está a cargo de dois digestores anaeróbios, os quais funcionam em paralelo, com um tempo mínimo de retenção total de 21 dias e uma carga de sólidos voláteis inferior a 2,5 Kg MV/m³.dia Cada digestor possui 11 m de diâmetro e 600 m³ de volume. Na parede do digestor, existe um medidor de pressão, que verifica o nível das lamas e concede a informação do volume disponível no digestor a qualquer momento. A agitação dentro do digestor realiza-se a partir da introdução de gás, com origem num feixe central de canas de injeção alimentado por compressores. O aquecimento de lamas é efetuado através do funcionamento de três grupos eletrobomba, que aspiram as lamas e fazem-nas passar por permutadores de calor;

- Após serem aquecidas, as lamas são em seguida reenviadas para o digestor e misturam-se com as lamas frescas. A alimentação de água quente advém de uma caldeira ou dos grupos de cogeração, funcionando ambos os grupos a biogás. A recirculação de água quente é feita através de três grupos eletrobomba;
- As lamas digeridas são enviadas para desidratação, com recurso a um parafuso excêntrico horizontal;
- O biogás produzido na digestão, antes de ser aproveitado como energia, atravessa um pote de purga de condensados à saída de cada digestor. Este pote assegura a retenção dos condensados, os quais são prejudiciais na etapa de aproveitamento do biogás;
- O sistema de aproveitamento de biogás existente na ETAR de Sesimbra é constituído por:
 - Um gasómetro esférico de dupla membrana e 100 m³ de volume, o qual está equipado com dois ventiladores, uma válvula de sobrepressão e uma sonda de nível que indica o nível de biogás no gasómetro de forma contínua;
 - Uma tocha de queima que recebe o biogás em excesso;
 - Uma caldeira de aquecimento de lamas que funciona com recurso a biogás como fonte de energia;
 - Um sistema de cogeração composto por dois motores, dois refrigeradores, radiadores, um dessulfurizador e dois compressores de biogás.

➤ Desidratação de lamas

- A etapa de desidratação de lamas é efetuada por uma centrífuga capaz de desidratar 210 kg de lamas por hora;
- As lamas desidratadas são elevadas por uma bomba transportadora até um dos dois silos existentes de 30 m³ de volume cada;
- A adição de polieletrólito sucede em dois períodos: o primeiro ocorre antes de as lamas serem desidratadas, através de duas bombas de parafuso excêntrico. Na tubagem de compressão da bomba transportadora, que conduz as lamas desidratadas para os silos de armazenamento, é também injetado polieletrólito para lubrificação da tubagem.

Serviços auxiliares (SIMARSUL, 2007):

➤ Linha de tratamento de odores

- O sistema de tratamento de odores instalado consiste numa lavagem química e uma biodesodorização;

- A lavagem química é composta pelos seguintes elementos:
 - Duas torres de lavagem de tipo vertical, com escoamento em contracorrente. Na primeira torre de lavagem, ocorre uma lavagem ácida, e na segunda torre efetua-se uma lavagem oxidante em meio alcalino;
 - Rede de aspiração que recolhe o ar poluído;
 - Descalcificador de água.
 - A biodesodorização necessita dos seguintes órgãos:
 - Rede de aspiração que recolhe o ar poluído;
 - Ventiladores;
 - Biofiltro.
- Poço de escorrências
- Recebe as escorrências do compactador de resíduos, classificador de areias, separador de gorduras, espessador, desidratação, purgas das torres de desodorização e das águas de lavagem do edifício;
 - Os caudais recolhidos são enviados de volta à cabeça da estação para tratamento.
- Cisterna de água tratada
- Localizada antes da descarga do canal de UV, armazena a água tratada que posteriormente será utilizada nas lavagens dos Biofor.
- Cisterna de água de serviço
- Localizada antes da descarga no emissário, armazena a água tratada que posteriormente será reutilizada na rede interna da ETAR.
- Grupo hidropneumático
- Grupo de funcionamento automático que eleva a água de serviço a toda a estação recorrendo a duas eletrobombas.
- Laboratório
- Laboratório onde são realizadas as análises comuns e necessárias para um bom funcionamento da ETAR.

Desenvolvimento do estudo

A metodologia empregue neste segundo caso de estudo é apresentada na Figura 3.8

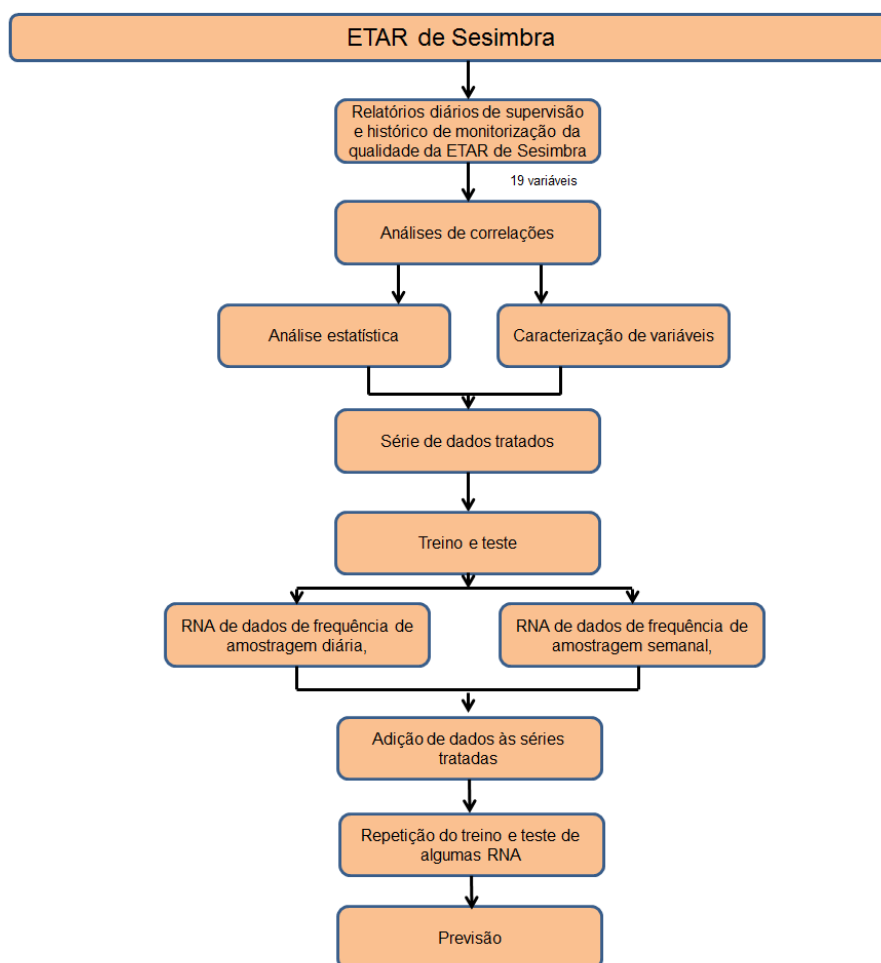


Figura 3.8 – Metodologia aplicada na investigação realizada na ETAR de Sesimbra

O segundo caso de estudo diz respeito à modelação do funcionamento da digestão anaeróbia da ETAR de Sesimbra, utilizando, tal como na ETAR do Seixal, as potencialidades das RNA e o programa de computador NeuralTools® como ferramenta de trabalho. Este estudo é semelhante à investigação realizada na ETAR do Seixal no ano de 2014, pelo que se tentou usar uma metodologia análoga, de modo a ser possível efetuar uma comparação dos resultados produzidos nos dois casos, em condições idênticas. No início da realização da dissertação, não existia certeza absoluta acerca da decisão de incluir ou não este caso de estudo na investigação a concretizar, algo que só foi deliberado quase a meio do período disponível para a execução da dissertação, pelo que a metodologia empregue apresenta as seguintes ressalvas:

- A metodologia selecionada neste caso de estudo foi desenvolvida de modo a ser parecida à empregue na ETAR do Seixal e a permitir a comparação dos resultados. No entanto, a

metodologia aplicada é mais simples, dado o tempo disponível para a sua execução ser inferior ao tempo de investigação já investido no primeiro caso de estudo;

- A quantidade de variáveis disponibilizadas para o estudo é inferior à da ETAR do Seixal e a quantidade de dados não apresenta a mesma duração temporal.

A primeira etapa do estudo passou pela verificação dos parâmetros com medições registadas e afetos ao processo de digestão anaeróbia. Após a constatação dos dados disponíveis ocorreu a recolha dos mesmos. A nível do registo de dados de carácter semanal ou com frequência de amostragem inferior a uma vez por semana, a recolha foi fácil e rápida, já que os dados se encontravam em formato digital. A nível do registo de dados de carácter diário, a sua recolha foi mais difícil e demorada, visto estes estarem disponíveis somente em suporte físico. À medida que a recolha ocorria, os dados eram acomodados em séries de dados.

Finalizada a recolha e a preparação de dados em séries, efetuou-se o cálculo do coeficiente de correlação de Pearson, geraram-se gráficos e realizou-se o tratamento estatístico das séries de dados, de modo a apurar-se a qualidade de dados disponíveis e o tratamento de dados a executar. Realizada a análise aos dados, realizaram-se alguns cortes nas séries de dados, de forma a que, na altura da produção das RNA, fosse possível testar diferentes séries de dados e comprovar as diferenças entre resultados.

O passo a tomar depois de concluída a preparação de dados foi a produção de RNA, com o objetivo final de estimar os valores de biogás gerados pelo digestor anaeróbio, a qual aconteceu de duas formas diferentes. A primeira forma de produção das RNA teve por base as séries de dados de frequência de amostragem diária, enquanto na segunda abordagem utilizam-se as séries de dados de índole semanal ou frequência de amostragem inferior a uma semana. No decorrer dos testes com as RNA, os resultados iam sendo observados e analisados. Este processo foi repetido numa fase posterior, dado que perto do final do período disponível para a execução dos testes, foram adicionados novos dados às séries. No final de todas as RNA terem sido geradas, procedeu-se à análise e tratamento dos resultados concebidos pela investigação efetuada neste caso de estudo.

As últimas ações a ser concretizadas foram uma comparação entre os coeficientes de correlação registados nos dois casos de estudo e a criação de uma rede neuronal com base em dados de ambos os casos de estudo. Por fim foi efetuada uma análise final de toda a investigação realizada.

4. NeuralTools®

Para o desenvolvimento de redes neurais artificiais é necessário o uso de uma ferramenta com a capacidade de as criar, gerindo o seu funcionamento com base nos dados disponíveis. O presente trabalho pretende dar continuidade a uma investigação previamente iniciada nas instalações da SIMARSUL, na qual se usou um programa informático denominado NeuralTools®, como motriz do estudo. A decisão de continuar a usar este mesmo programa foi lógica, dado tratar-se de uma ferramenta de aplicação que já provou ser capaz de produzir resultados interessantes, e com a qual, os engenheiros responsáveis pelo funcionamento das instalações afetadas ao caso de estudo, estão familiarizados. O processo de análise e comparação dos resultados produzidos no passado face aos do presente também se torna muito mais fácil quando se usa o mesmo instrumento de trabalho. A única diferença verificada diz respeito à versão do programa escolhida. Enquanto, no ano passado, se utilizou a versão 5.5.0 do NeuralTools®, no atual estudo, usou-se a versão 6.

O NeuralTools® é um programa de computador concebido pela Palisade®, uma das empresas líderes na criação de programas de análise de risco e tomada de decisões, a nível mundial. Perante a ausência de informação atualizada, pode ser um grande auxiliar na tomada de decisões, caso existam dados antigos análogos à informação em falta, uma vez que se trata de um programa sofisticado em prospeção de dados, capaz de fazer predições precisas, baseadas em padrões existentes em informação antiga. Os princípios básicos do seu funcionamento são semelhantes aos de funções cerebrais, o que permite resolver problemas a nível categórico e numérico. Esta ferramenta informática pode ser e é usada em diversas áreas profissionais, desde a medicina à arqueologia (Palisade, 2015).

As RNA podem aparentemente ser difíceis de criar e de compreender, mas através do uso do NeuralTools®, esta tarefa torna-se mais fácil. A principal característica e vantagem deste programa centra-se no facto de permitir desenvolver o trabalho numa plataforma familiar a qualquer engenheiro: o Microsoft Excel®. Funciona como um suplemento, que apenas adiciona uma barra de ferramentas à interface do Excel, tornando o seu manuseamento bastante simples e intuitivo. Em seguida, explicar-se-ão detalhadamente as funcionalidades e potencialidades do NeuralTools®.

4.1 Funcionamento do NeuralTools®

Concluída a instalação no computador, ao iniciar o programa temos acesso à barra de ferramentas do NeuralTools® no Microsoft Excel®, como é visível na Figura 4.1.

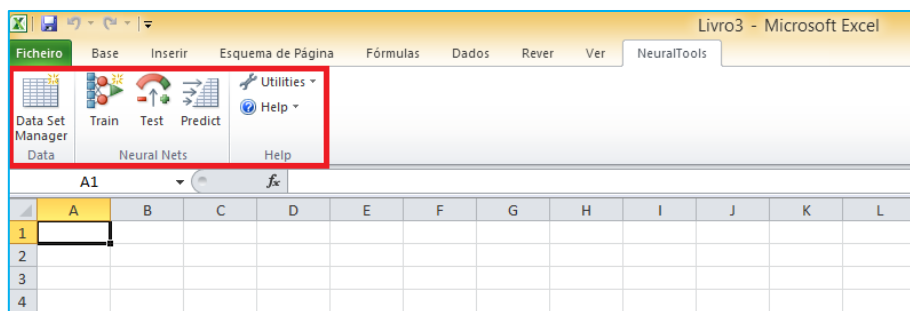


Figura 4.1 – Localização da barra de ferramentas do NeuralTools® na interface do Microsoft Excel®

O modo de funcionamento do programa está dividido em quatro fases: a primeira corresponde à preparação de dados (comando Data Set Manager); a segunda ao treino da rede (comando Train); a terceira ao período de teste (comando Test); e a última fase diz respeito à previsão (comando Predict). Todas estas etapas têm um comando presente na barra de ferramentas do NeuralTools®, na qual ainda é possível aceder a outras funções úteis na produção de uma RNA (comando Utilities) e a um suporte de ajuda (Help). A barra de ferramentas do NeuralTools® pode ser observada na Figura 4.2.

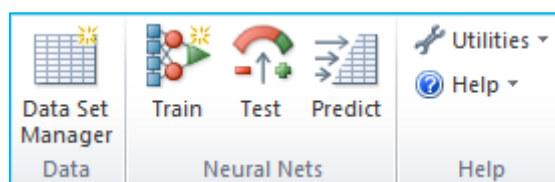


Figura 4.2 – Barra de ferramentas do NeuralTools®, onde estão inseridos os comandos de controlo do programa

A primeira fase da produção de uma rede neuronal artificial através da utilização do NeuralTools® é a preparação de dados. Geralmente é uma etapa importante na produção de uma rede neuronal artificial, visto em muitas situações antes de se usar o comando *Data Set Manager*, ser necessário passar algum tempo a trabalhar os dados. Os resultados só serão satisfatórios se o conjunto de dados formado por variáveis e casos conceder informação de qualidade e estiver bem organizado. As variáveis (exemplo de um conjunto de valores de uma variável a preto na Figura 4.3) possuem os seus valores alinhados em colunas contíguas, correspondendo cada linha do conjunto de dados a um caso (a vermelho na Figura 4.3), formado por um conjunto de variáveis independentes, a partir do qual se quer prever um valor para a variável dependente.

Data Set Manager Data		Train	Test	Predict	Utilities ▾ Help ▾
G9		fx			
	A	B	C	D	E
1	Data	Q lamas (m3/dia)	TRH (dias)	T reator (°C)	Q biogás (m3/dia)
29	28/01/2013	13,06	45,9	31,17	326
30	29/01/2013	13,14	45,7	31,48	322
31	30/01/2013	12,80	46,9	30,69	263
32	31/01/2013	12,35	48,6	32,69	185
33	01/02/2013	13,01	46,1	30,00	847

Figura 4.3 – Estrutura de dados, formada por variáveis e casos

Cada variável está associada a um nome e a um intervalo de células do Excel. O comando *Data Set Manager*, exibido na Figura 4.4, permite selecionar, editar ou apagar o conjunto de dados em utilização, ao definir quais as variáveis ativas e a extensão de células a considerar. Quando tentamos, através deste comando, definir um conjunto de dados, o NeuralTools® automaticamente procura definir as variáveis e a extensão de células, tornando este processo mais rápido (Palisade, 2013).

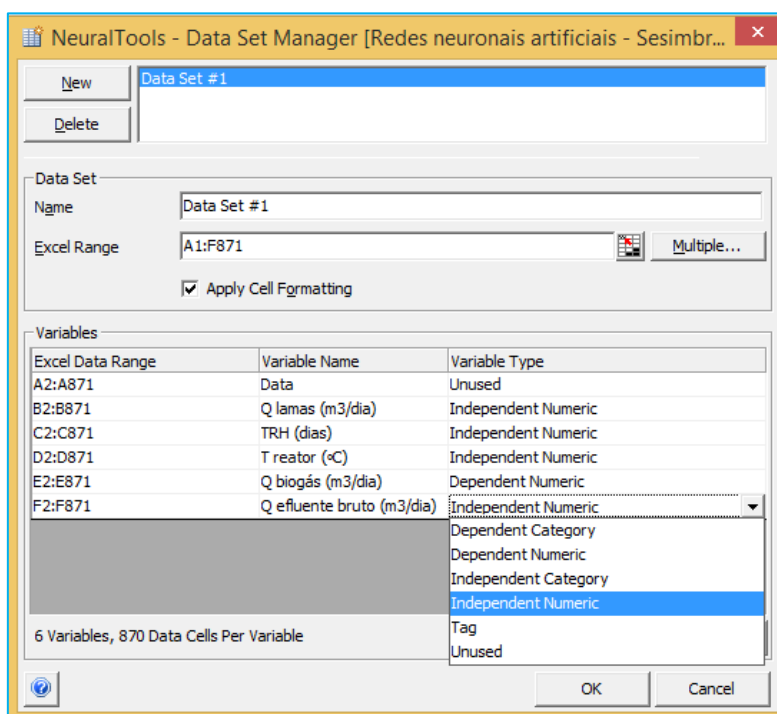


Figura 4.4 – Caixa de diálogo do comando Data Set Manager

Na caixa de diálogo do comando *Data Set Manager* estão disponíveis as seguintes opções (Palisade, 2013):

- Adicionar ou apagar um conjunto de dados;
- Especificar o nome do conjunto de dados;
- Alterar o intervalo de células que forma o conjunto de dados, através do botão *Multiple*, o qual permite a seleção de múltiplos intervalos, ficando ao critério do utilizador a sua eleição;
- Formatar as células ao colorir e definir limites para o conjunto de dados;
- Definir o tipo de variáveis que formam o conjunto de dados. É possível classificar uma variável como categórica ou numérica, independente ou dependente. No caso da existência de variáveis categóricas, a rede é treinada com o intuito de, a partir da informação das variáveis independentes, classificar a variável dependente numa categoria como “azul” ou “vermelho”. A nível das variáveis numéricas, a rede é treinada com o objetivo de, a partir dos dados das variáveis independentes, predizer um valor numérico para a variável dependente. É ainda possível classificar uma variável como “Tag”, a qual é usada para identificar casos que serão usados especificamente para treino, teste ou previsão. Por fim, existe a classificação “não usada”, que representa a não utilização da variável em questão para a génese da rede neuronal;
- Importar tipos de variáveis de um conjunto de dados para outro e vice-versa, treinadas ou não.

Uma vez definidas as opções no comando *Data Set Manager* relativamente ao conjunto de dados, é possível avançar para as etapas de treino, teste e previsão da rede neuronal. Dependendo da natureza e qualidade de dados a usar, geram-se várias possibilidades de criação de redes neuronais. As várias redes treinadas apresentam capacidades de predição diferentes: umas serão mais precisas, outras serão de menor qualidade. O processo de testes, após o treino, esclarece-nos acerca do quão boa será a capacidade de previsão da rede gerada. No fim do treino e teste da rede, é possível executar a previsão de valores, embora, na maior parte das vezes, esta etapa só valha a pena ser executada se as fases anteriores apresentarem bons resultados.

A etapa referente ao treino pode ocorrer de forma isolada, em simultâneo com a etapa de teste ou em conjunto com as etapas de treino e previsão. No presente trabalho, optou-se por acoplar as fases de treino e de teste, tendo a fase de previsão sido realizada posteriormente, apenas nas redes que demonstraram um maior potencial. Na caixa de diálogo do comando *Train*, a qual pode ser visualizada na Figura 4.5, permite definir várias opções relativas à etapa de treino.

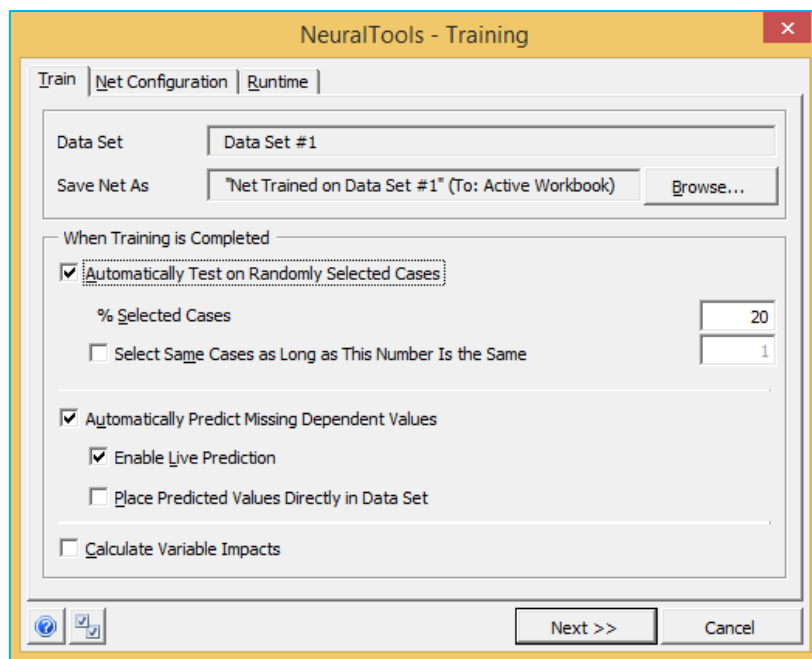


Figura 4.5 – Caixa de diálogo do comando Train

A secção da caixa de diálogo relativa ao treino permite visualizar ou tomar as seguintes opções (Palisade, 2013):

- Indica o conjunto de dados disponível para iniciar o treino de rede;
- Mudar a localização onde a rede será guardada;
- Concede a opção de treinar e testar em simultâneo a rede. Caso se opte por efetuar a fase de testes logo após a fase de treino, é necessário seleccionar a percentagem de casos aleatórios que se quer deixar para teste, os quais não farão parte do treino. Casos identificados para teste por uma variável seleccionada como “Tag” no *Data Set Manager* também não entram na fase de treino, apenas na fase de testes;
- Ativar a previsão automática dos valores dependentes em falta. É possível prever os casos em que o valor da variável dependente está em falta e os casos identificados para previsão, por uma variável “Tag” no *Data Set Manager*;
- Concede permissão ao programa para estimar os valores existentes em cada caso da variável dependente, a partir de uma fórmula criada que não tem em conta o seu valor real;
- Calcular o impacto das variáveis, ou seja, a influência que cada variável independente tem na fase de treino e que posteriormente terá na previsão de valores por parte da rede.

O NeuralTools® pode desenvolver diferentes configurações de redes, testando várias hipóteses, de modo a gerar uma rede com a melhor capacidade de previsão possível. A secção relativa à configuração de rede na caixa de diálogo do comando *Train*, exibida na Figura 4.6, permite escolher qual ou quais os tipos de redes que queremos que o programa crie.

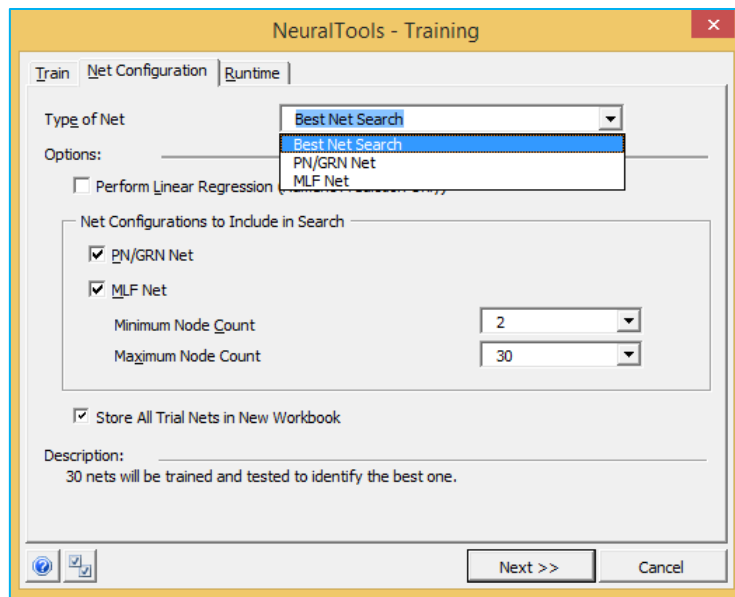


Figura 4.6 – Secção referente à configuração de rede do comando Train

Na secção de configuração de rede é possível fazer as seguintes escolhas (Palisade, 2013):

- Executar uma regressão linear. Acionar esta opção inibe o cálculo do impacto das variáveis;
- Escolher o tipo de redes a desenvolver. Se as variáveis do conjunto de dados ativo estiverem selecionadas como categóricas, é permitido gerar redes neuronais do tipo probabilístico (PNN) e do tipo *Multi-layer Feedforward* (MLFN). No caso de estarem em utilização variáveis numéricas, as redes possíveis de conceber são as MLFN e as de regressão generalizada (GRNN), idênticas às PNN. Existe a opção de selecionar um tipo de rede específico ou ativar a procura pela melhor rede. Neste caso, o programa irá testar vários tipos de redes e identificar a melhor rede produzida. Só é possível acionar esta opção ao estipular que a etapa de teste ocorrerá posteriormente á fase de treino, dado a melhor rede ser definida com base nos resultados do período de testes. Tomando a opção pela procura da melhor rede, é necessário definir o intervalo de nós mínimo e máximo a considerar para as redes do tipo MLFN;
- Guardar todas as redes produzidas, o que permite mais tarde usar cada uma delas na etapa de previsão. Esta opção só está disponível se a procura pela melhor rede estiver ativa.

No atual estudo, as variáveis usadas pertenceram sempre ao domínio das numéricas, pelo que as redes desenvolvidas foram as GRNN e as MLFN. A opção relativa à procura pela melhor rede foi continuamente selecionada, tal como se optou sempre por guardar todas as redes produzidas.

Por fim, no comando *Train*, é essencial selecionar as condições de paragem do treino em uma caixa de diálogo que pode ser observada na Figura 4.7. Existem três hipóteses de escolha para

parar o treino, que podem funcionar isoladamente ou em conjunto. Ao se optar por combinar as três condições, o treino será interrompido assim que for satisfeita uma condição. Se não forem tomadas quaisquer opções, o treino acabará eventualmente, sendo que o período de tempo relativo ao tipo de redes GRNN é muito inferior ao das MFLN (Palisade, 2013).

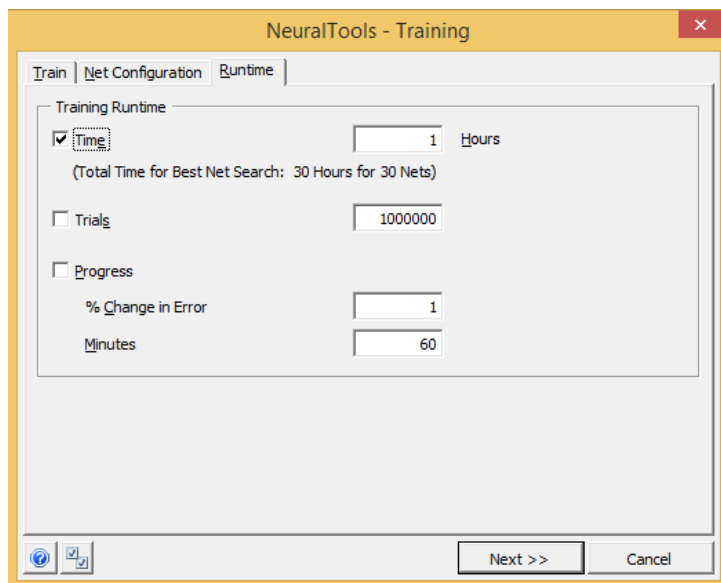


Figura 4.7 – Secção referente às condições de paragem de treino, do comando Train

Na secção afeta às condições de paragem de treino, podem seleccionar-se as seguintes opções (Palisade, 2013):

- Especificar um tempo, que faz cessar o período de treino após ser atingido. O treino pode parar antes de atingir o tempo definido, caso o algoritmo determine que é improvável que continuem a existir progressos. Se a procura pela melhor rede estiver ativa, o tempo estipulado será cumprido obrigatoriamente, de forma a não permitir que o programa se dedique mais a uma rede do que a outra;
- Escolher o número de tentativas máximo que o programa pode fazer antes de o treino parar;
- Selecionar a percentagem mínima de melhoria de erros que deve existir num intervalo de tempo estipulado para impedir a paragem do treinamento.

Uma vez configuradas todas as opções relativas ao treinamento, o NeuralTools® exibe um menu, suscetível de ser observado na Figura 4.8, onde é possível arrancar com a fase de treino e visualizar um sumário das condições em que este decorrerá.

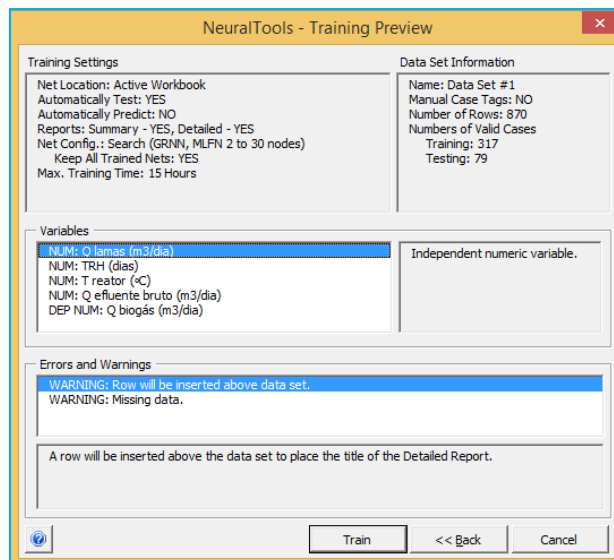


Figura 4.8 – Caixa de diálogo onde são exibidas as condições em que o treino decorrerá

Visualizando o sumário das condições de treino, é possível rever se as definições seleccionadas vão de encontro à intenção do utilizador. O programa tem ainda a particularidade de identificar possíveis problemas ou erros no conjunto de dados, oferecendo ao utilizador a possibilidade de os corrigir antes de iniciar o treino de rede.

Durante a fase de treino, temos acesso a uma janela, representada na Figura 4.9, onde é reportado o desempenho de cada rede à medida que esta é treinada. Existe também a opção de fazer uma pausa ou de parar totalmente o processo de treino, seja de uma rede ou de todas.

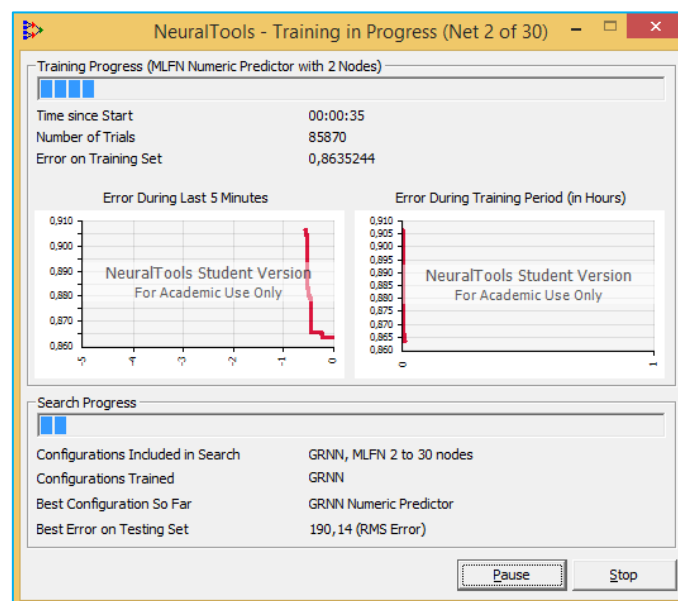


Figura 4.9 – Caixa de diálogo, onde é possível acompanhar o processo de treino das redes neuronais

Os gráficos exibidos indicam as melhorias ocorridas na redução de erros da rede em treino, embora não se deva utilizar esta informação para inferir acerca da qualidade de previsão da rede. Acima do gráfico, é visível a duração do treino da rede em causa, o número de ensaios já realizados e o erro médio quadrático dos dados treinados, com base numa escala selecionada durante o treino. Perto do limite inferior da janela, constata-se várias informações, nomeadamente, o nome das redes incluídas na etapa de treino, as redes treinadas até ao momento, a rede que evidenciou maior qualidade na fase de testes até ao instante e o erro médio quadrático que lhe é associado (Palisade, 2013).

No fim da fase de treinamento, é gerado um relatório detalhado, com base em dados estatísticos e gráficos, da melhor rede criada. Elegida a opção de guardar todas as redes, existente no comando *Train*, é também produzido um relatório detalhado de cada rede concebida. O relatório nem sempre é igual, pois depende das opções selecionadas em passos anteriores e do facto de as fases de teste e previsão estarem ou não acopladas à fase de treino. No Anexo I, é possível consultar um exemplo de um relatório detalhado da fase de treino e teste. As principais informações para previsões numéricas que podem ser consultadas num relatório de treino são as seguintes (Palisade, 2013):

- Taxa de más previsões ocorridas durante o treino e teste. Uma má previsão é identificada quando a previsão falha o valor real, por uma margem definida, que normalmente é de cerca de 30%;
- Erro médio quadrático verificado durante o treino e teste, que representa o desvio do valor das previsões dos valores reais;
- Erro médio absoluto registado durante o treino e teste, que representa o desvio médio do valor das previsões dos valores reais;
- Impacto de cada variável independente nos resultados gerados;
- Histogramas residuais, onde se pode observar a diferença entre o valor real e o previsto;
- Gráficos de dispersão que mostram a relação entre os valores reais, os previstos e os residuais.

Finalizados o treino e os testes, a fase seguinte é a previsão, onde uma rede treinada normalmente irá estimar valores correspondentes a novos casos, ou seja, casos em que a rede não conhece o valor da variável dependente. Ao clicar-se sobre o botão *Predict* na barra de ferramentas do NeuralTools®, tem-se acesso às definições da etapa de previsão. A caixa de diálogo do comando *Predict* pode ser visualizada na Figura 4.10.

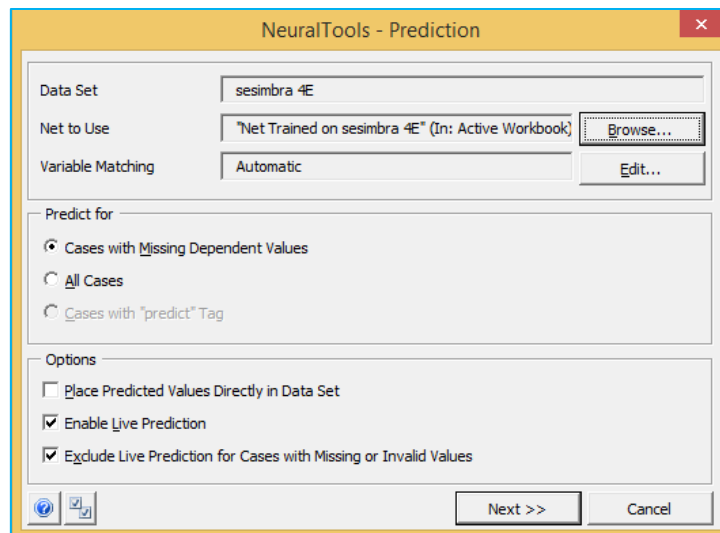


Figura 4.10 – Caixa de diálogo do comando Predict

Na caixa de diálogo do comando *Predict*, é possível observar e configurar as seguintes definições (Palisade, 2013):

- Visualizar o conjunto de dados que será usado para previsão, o qual necessita de ser selecionado no *Data Set Manager* e estar ativo;
- Especificar o nome e a localização da rede a usar na previsão;
- Fazer a correspondência entre as variáveis do conjunto de dados usados na previsão e as variáveis do conjunto de dados treinados;
- Selecionar os casos para os quais irá ser realizada a previsão. Por norma, executa-se a predição para casos em que o valor da variável dependente é desconhecido, mas também é possível para os casos em que o valor é conhecido. Se anteriormente foi eleita uma variável como “Tag”, a previsão realizar-se-á apenas para os casos marcados com esse intuito por este tipo de variável;
- Colocar os valores previstos diretamente no conjunto de dados, o que sobrescreve qualquer valor anterior. Apesar de ser possível identificar os valores alterados, devido ao facto de possuírem uma cor diferente, esta opção deve ser usada com cuidado;
- Ativar a previsão ao vivo, a qual é uma ferramenta bastante característica e útil do NeuralTools®. Permite coligar fórmulas às células análogas aos valores previstos da variável dependente. Assim que os valores independentes são alterados, o valor da variável dependente é recalculado. Existe a opção de ativar a previsão ao vivo, mas de não permitir que esta calcule valores para os casos com valores inválidos ou em falta.

Definidas as configurações, é exibida uma nova caixa de diálogo que pode ser observada na Figura 4.11, onde é possível observar um resumo das escolhas realizadas e os erros detetados pelo programa. Este resumo é útil, dado permitir comprovar se está tudo de acordo com o que o utilizador pretende e verificar se é necessário corrigir algum lapso.

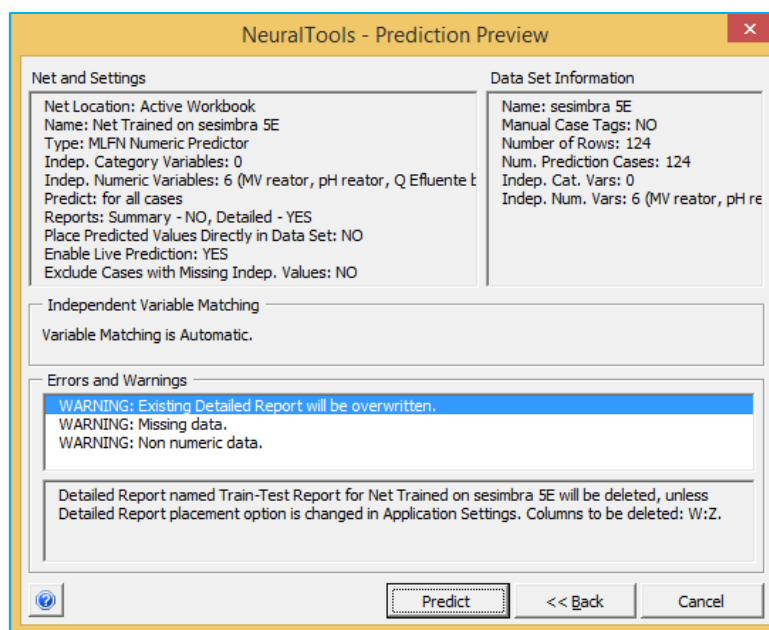


Figura 4.11 – Caixa de diálogo com o resumo das definições da etapa de previsão

Ao clicar-se sobre o botão *Predict*, arranca a fase de previsão, que processa de forma bastante mais rápida que a fase de treino e teste. Os resultados podem ser colocados diretamente nas células correspondentes ao valor da variável dependente, ou dispostos num relatório detalhado localizado em colunas próximas do conjunto de dados.

O NeuralTools® possui ainda outro comando importante, para além dos que dão acesso às fases de desenvolvimento de uma RNA. O comando *Utilities* (passível de ser observado na Figura 4.12) foi desenvolvido para auxiliar o processo de criação e modelação de uma rede, através dos seus subcomandos (Palisade, 2013):

- “Configuração de aplicações” é um subcomando que permite alterar algumas definições padrão das fases de treino, teste e previsão;
- O subcomando “gerente de rede neuronal” concede que se transfiram redes neuronais treinadas, entre pastas de trabalho ou arquivos, às quais se pode adicionar informação descritiva;
- A partir do subcomando “falta de dados”, é possível substituir os dados em falta ou com erro por valores artificiais estatísticos;
- O subcomando “teste de sensibilidade” ajuda a determinar o grau de fiabilidade dos resultados da fase de teste, aferindo se a quantidade de casos destacados para teste é significativa ou necessita de aumentar.

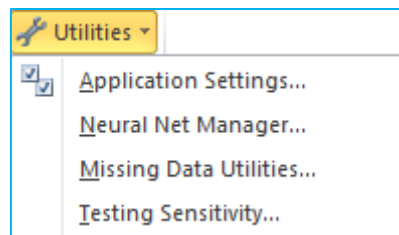


Figura 4.12 – Subcomandos do comando Utilities

5. Resultados e Discussão

5.1 Caso de Estudo – ETAR do Seixal

A primeira vertente do trabalho passou por recolher e preparar os conjuntos de dados utilizados na investigação realizada no ano passado, construir novas séries de dados e reproduzir algumas das redes neuronais criadas no ano passado (Pinto, 2014) A segunda vertente da investigação diz respeito à análise da influência na digestão anaeróbia, da ocorrência de intrusões salinas no efluente bruto que chega à ETAR do Seixal.

5.1.1 Recolha de Dados e Construção das Novas Séries de Dados

Do estudo executado em 2014, foram recolhidos para o presente trabalho dados de cerca de 20 variáveis afetas ao funcionamento do digestor anaeróbio, as quais estão expostas na Tabela 5.1. A escolha recaiu sobre este grupo, dado ter sido a partir das séries de dados destas variáveis que se produziram todas as RNA no primeiro ano de investigação na ETAR do Seixal.

Tabela 5.1 – Variáveis transpostas para a presente investigação

Variável	Unidade	Frequência de amostragem	Número de dados disponíveis	Forma de obtenção
Q biogás	m ³ /dia	diária	854	medição
Q lamas	m ³ /dia	diária	880	medição
T reator	oC	diária	823	medição
TRH	dias	diária	880	cálculo
AGV reator	mg/L	semanal	95	análise laboratorial
AGV/TAL	-	semanal	95	cálculo
CQO in	mg/L	semanal	78	análise laboratorial
CQO reator	mg/L	semanal	82	análise laboratorial
CV	kg/(m ³ .dia)	semanal	159	cálculo
LCQO in	kg/dia	semanal	71	cálculo
LCQO out	kg/dia	semanal	75	cálculo
LCQO removida	kg/dia	semanal	71	cálculo
LMV in	kg/dia	semanal	159	cálculo
LMV out	kg/dia	semanal	159	cálculo
LMV removida	kg/dia	semanal	159	cálculo
MV in	mg/L	semanal	167	análise laboratorial
MV reator	mg/L	semanal	167	análise laboratorial
pH in	-	semanal	136	análise laboratorial
pH reator	-	semanal	96	análise laboratorial
TAL reator	mg/L	semanal	100	análise laboratorial

Às séries de dados recolhidas, foram adicionados novos dados. O número de dados adicionados por variável e a quantidade de dados totais disponíveis para a produção de RNA, podem ser consultados na Tabela 5.2.

Tabela 5.2 – Número de dados disponíveis para a produção de RNA

Variável	Unidade	Número de dados disponíveis na investigação anterior	Número de dados adicionados	Número de dados totais disponíveis
Q biogás	m ³ /dia	854	281	1135
Q lamas	m ³ /dia	880	281	1161
T reator	°C	823	281	1104
TRH	dias	880	281	1161
AGV reator	mg/L	95	35	130
AGV/TAL	-	95	35	130
CQO in	mg/L	78	19	97
CQO reator	mg/L	82	19	101
CV	kg/(m ³ .dia)	159	34	193
LCQO in	kg/dia	71	19	90
LCQO out	kg/dia	75	19	94
LCQO removida	kg/dia	71	19	90
LMV in	kg/dia	159	34	193
LMV out	kg/dia	159	34	193
LMV removida	kg/dia	159	34	193
MV in	mg/L	167	36	203
MV reator	mg/L	167	36	203
pH in	-	136	35	171
pH reator	-	96	36	132
TAL reator	mg/L	100	35	135

A localização de onde foram retiradas as amostras para análises laboratoriais e o local onde foram medidos diretamente os valores de algumas variáveis podem ser consultados no Anexo II. Os respetivos gráficos das novas séries de dados de frequência de amostragem diária são apresentados de seguida (Figura 5.1, Figura 5.2, Figura 5.3 e Figura 5.4) Os gráficos das restantes variáveis são apresentados no Anexo III.

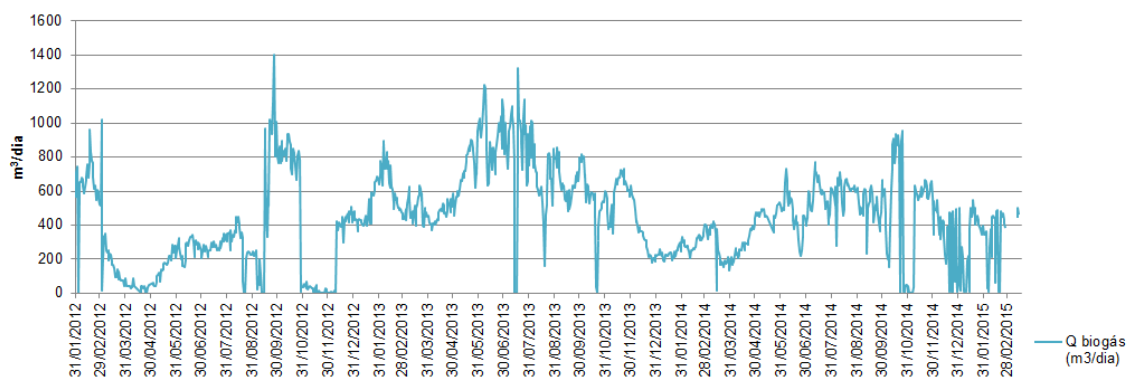


Figura 5.1 – Distribuição temporal da variável “Q biogás” – ETAR do Seixal

Na Figura 5.1 é possível constatar que a série de dados de caudal de biogás é irregular, apresentando grande variância nos valores registados.

A informação transmitida pela Figura 5.2 permite confirmar que de facto a série de dados do caudal de biogás produzido não apresenta uma distribuição normal dos valores registados. A Figura 5.2 indica tal como a Figura 5.1 que existe elevada variabilidade nos valores registados, no que diz respeito à variável “Q biogás”.

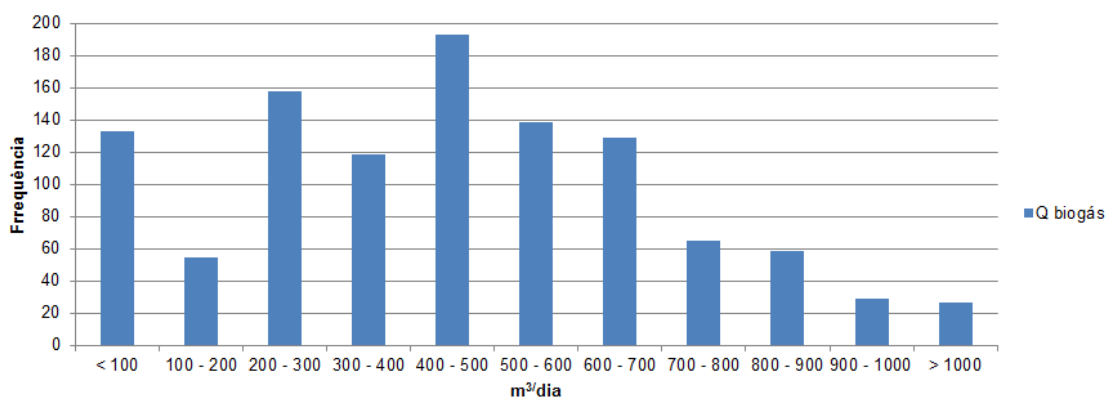


Figura 5.2 – Frequência de valores registados da série de dados da variável “Q biogás” – ETAR do Seixal

A partir da Figura 5.3 é possível comparar os valores registados das variáveis “Q biogás” e “Q lamas”, permitindo observar que parece não existir grande correlação entre os valores registados de cada variável, dado que a olho nu é difícil identificar qualquer padrão nos dados registados.

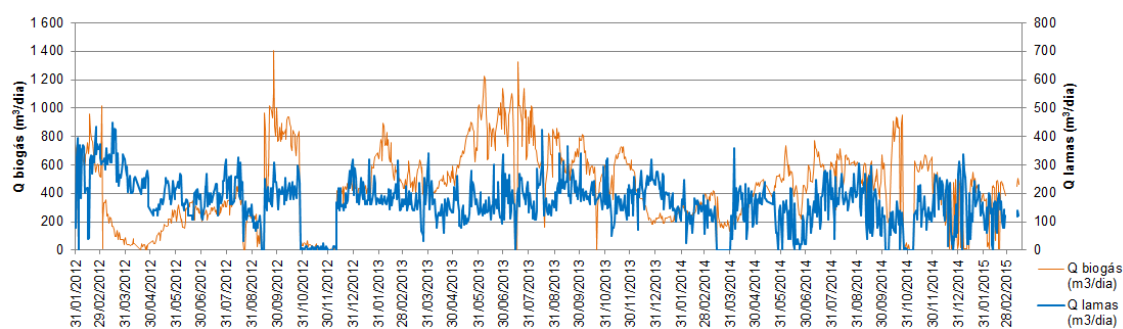


Figura 5.3 – Distribuição temporal das variáveis “Q biogás” e “Q lamas” – ETAR do Seixal

Do mesmo modo, através da análise da Figura 5.4 constata-se que a temperatura registada no digestor anaeróbio é mais elevada nos meses normalmente mais quentes do ano em Portugal e mais baixa na estação mais fria. Em relação aos valores do caudal de biogás produzido, estes parecem ser mais elevados quando a temperatura no digestor é também mais elevada, muito embora exista alguma irregularidade na série de valores do caudal de biogás, pelo que esta relação não é linear.

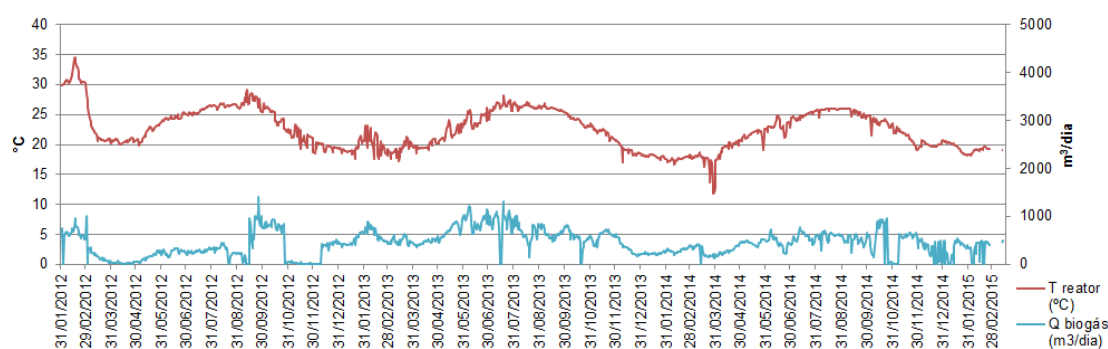


Figura 5.4 – Distribuição temporal das variáveis “Q biogás” e “T reator” – ETAR do Seixal

Com base na variação de temperatura verificada no digestor, quis-se perceber se a média de valores registados das séries de dados de frequência com amostragem diária indicavam algum efeito de sazonalidade. A média de valores das séries de dados pode ser consultada na Tabela 5.3.

Tabela 5.3 – Média dos valores registados afetos às variáveis “Q lamas”, “T reator” e “Q biogás” – ETAR do Seixal

	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Q lamas	193,64	196,59	215,92	184,11	171,32	167,64	189,55	186,83	198,86	187,78	154,61	208,60
T reator	19,03	21,93	20,28	19,97	21,97	24,25	25,65	26,26	26,20	24,38	21,66	19,53
Q biogás	354,85	487,13	363,59	284,23	453,64	583,68	573,15	529,40	594,02	693,19	596,40	366,82

A análise da Tabela 5.3 confirma que temperatura do digestor é claramente mais elevada nos meses mais quentes do ano, parecendo influenciar a produção de biogás, uma vez que a média mensal dos valores de biogás produzido é mais elevada quando a temperatura no

digestor excede os 21°C na média mensal. No que diz respeito ao caudal de lamas espessadas, este não indica variações sazonais, confirmando-se inclusive que a correlação entre esta variável e o caudal de biogás parece ser realmente fraca.

No global, os registos de dados utilizados datam de 31 de Janeiro de 2012 a 17 de Março de 2015. Os resultados do cálculo do coeficiente de correlação de Pearson das novas séries de dados, podem ser observados na Tabela 5.4. As séries de dados utilizadas para o cálculo deste coeficiente não englobam valores igual a zero e no caso da variável “Q lamas”, apenas foram considerados os valores entre os 75 m³ e os 375 m⁵. Esta consideração aliada ao facto de a variável “TRH” ser uma função direta dos valores de caudal de lamas espessadas, levou a que não se apresentasse qualquer análise gráfica desta variável. A média de valores desta série de dados apresenta o valor de 11,58 dias, encontrando-se dentro do limite inferior normal para uma digestão em alta carga. O critério de seleção em relação à variável “Q lamas” foi aplicado, tendo em conta a sua utilização no trabalho de 2014, o qual não foi alterado devido a duas razões:

- O critério de aplicação é considerado viável. Valores iguais ou próximos de zero indicam problemas de operação, os quais, se forem considerados, fogem aos padrões de funcionamento normal do digestor, o que causará dificuldades na aprendizagem das RNA. O caudal de projeto de lamas afluentes ao digestor anaeróbio é de 188 m³ por dia no ano zero e de 253 m³ por dia no ano horizonte, pelo que, valores fora do critério selecionado afastam-se bastante da afluência normal de lamas. Se fossem considerados valores exteriores à regra aplicada, mais uma vez poderia impossibilitar a produção de resultados viáveis pelas RNA;
- Ao utilizar-se o mesmo critério, a comparação é mais fácil, tornando mais acessível a compreensão do impacto dos valores adicionados às séries de dados.

A análise dos resultados do cálculo do coeficiente de Pearson é efetuada segundo os seguintes critérios:

- Os valores de r entre zero e 0,30 significam que existe uma correlação linear fraca entre as variáveis. Se o resultado de r se encontrar entre 0,30 e 0,70, a correlação linear entre variáveis é moderada (valores a amarelo na Tabela 5.4), enquanto valores superiores a 0,70 indicam que existe uma correlação linear forte (valores a verde na Tabela 5.4).

Tabela 5.4 – Resultados do cálculo do coeficiente de correlação de Pearson das séries de dados afetas ao digestor anaeróbio da ETAR do Seixal

	Q biogás	Q lamas	T reator	TRH	AGV reator	AGV/TAL	CQO in	CQO reator	CV	LCQO in	LCQO out	LCQO removida	LMV in	LMV out	LMV removida	MV in	MV reator	pH in	pH reator	TAL reator
Q biogás																				
Q lamas	-0,02																			
T reator	0,42	0,21																		
TRH	0,02	-0,91	-0,15																	
AGV reator	-0,44	-0,22	-0,02	0,12																
AGV/TAL	-0,45	-0,08	-0,07	0,25	0,78															
CQO in	0,01	-0,19	-0,26	0,19	0,13	-														
CQO reator	0,01	-0,24	-0,11	0,13	0,18	-	0,33													
CV	0,35	-	-	-	-	-	-	-												
LCQO in	0,16	0,40	-0,24	-	-0,04	-	0,74	0,23	-											
LCQO out	0,15	0,61	-0,14	-	-0,21	-	-	0,62	-	0,52										
LCQO removida	0,04	-0,02	-0,07	-	-0,14	-	-	-0,07	-	0,74	-0,06									
LMV in	0,35	0,57	-0,16	-0,39	-0,17	-	-	0,11	1,00	0,50	0,45	0,14								
LMV out	-0,04	0,71	-0,05	-0,56	-0,01	-	-	0,01	-	0,39	0,53	0,11	0,57							
LMV removida	0,50	0,19	-0,06	-0,15	-0,18	-	-	0,06	-	0,11	-0,04	0,17	0,72	0,11						
MV in	0,39	-0,32	-0,30	0,27	-0,05	-		0,38	-	-	0,06	0,18	0,50	-0,10	0,52					
MV reator	-0,09	-0,22	-0,08	0,07	0,33	0,13	0,24	0,47	-	0,20	0,22	0,02	0,07	0,46	-0,14	0,32				
pH in	0,46	0,18	0,28	-0,05	-0,55	-0,06	-	-	-	-0,20	-	-0,25	0,02	-	0,11	-0,07	-			
pH reator	0,43	-0,04	-0,01	-0,05	-0,22	-0,44	-	0,06	-	0,13	0,11	0,12	0,01	-0,26	0,19	0,04	-0,31	0,16		
TAL reator	-0,06	-0,35	0,12	0,04	0,38	-0,47	0,15	0,20	-	0,15	0,01	0,15	-0,15	-0,09	-0,04	0,12	0,32	-0,28	0,33	

Face aos resultados do cálculo do coeficiente de correlação de Pearson das séries de dados antigas (resultados disponíveis no Anexo IV), no que diz respeito a estas vinte variáveis, o número de correlações lineares fortes entre variáveis aumentou. Nas séries antigas registavam-se correlações fortes entre os dados das variáveis “CV” e “LMV in”; “AGV reator” e “pH in”; “TRH” e “Q lamas”. Nas novas séries de dados existe uma correlação forte entre os dados das variáveis “CV” e “LMV in”; “TRH” e “Q lamas”; “AGV/TAL” e “AGV reator”; “LCQO removida” e “LCQO in”; “LMV removida” e “LMV in”. Este aumento no número de correlações fortes parece indicar que os dados adicionados às séries tiveram um impacto positivo nas correlações de algumas variáveis, no entanto também se registam algumas deduções nos valores de correlação entre outras variáveis.

A nível da correlação linear entre os dados da variável “Q biogás” e os dados das restantes variáveis, existem algumas alterações. No cálculo efetuado em 2014, o caudal de biogás apresenta uma correlação linear moderada com dez variáveis, que são as seguintes: “AGV reator”; “CV”; “LMV in”; “LMV removida”; “MV in”; “pH in”; “pH reator”; “Q lamas”; “T reator”; “TRH”. Nas séries construídas este ano, a série de dados da variável “Q biogás” exhibe uma correlação moderada com as seguintes nove variáveis: “AGV reator”; “AGV/TAL”; “CV”; “LMV in”; “LMV removida”; “MV in”; “pH in”; “pH reator”; “T reator”. No geral, existe um decréscimo da força de correlação linear entre a variável Q biogás e as restantes variáveis, sendo necessário destacar a redução de 0,44 para 0,02 entre a correlação linear do caudal de biogás e o caudal lamas elevadas. Este decréscimo registado, leva a crer que se poderá ter aplicado outros critérios de seleção de dados, além dos já enunciados e usados no trabalho deste ano. No cômputo global, os resultados do cálculo do coeficiente de Pearson realizado neste estudo, indicam que os dados adicionados na presente investigação deverão apresentar um nível qualitativo semelhante aos dados já existentes.

5.1.2 Previsão de Valores de Produção de Biogás

Os resultados da investigação concretizada no ano de 2014 permitiram detetar cinco RNA com um potencial de previsão de valores de produção de biogás superior às restantes redes neuronais geradas. As características e resultados das cinco melhores RNA produzidas no trabalho anterior, no que diz respeito à etapa de treino e teste, podem ser consultadas na Tabela 5.5, onde é possível verificar as seguintes condições (as quais são iguais para todas as restantes Tabelas deste tipo incluídas no restante trabalho):

- Nome da rede neuronal;
- As variáveis que deram origem a cada rede e suas respetivas unidades;
- A influência (impacto) de cada variável nos resultados gerados;
- O número de casos utilizados na fase de treinamento e teste de cada rede neuronal;
- A data inicial e final do período temporal do conjunto de dados utilizados;

- O tipo de rede neuronal que gerou melhores resultados, bem como a taxa de más previsões na fase de treino e teste e o erro médio quadrático associado a cada fase.

Tabela 5.5 – Cinco melhores redes neuronais produzidas no trabalho de 2014 (adaptado de Pinto, 2014)

Rede	Variáveis entrada	Impacto	Nº casos	Data		Resultados	
8-F	AGV reator (mg/L)	25,84%	42	20/03/2012	06/05/2014	Melhor rede	GRNN
	TAL reator (mg/L)	24,45%				% más previsões (treino - 30% toler.)	2,94%
	pH reator	23,05%				% más previsões (teste - 30% toler.)	37,50%
	T reator (°C)	17,54%				Erro médio quadrático (treino)	45,05
	TRH (dias)	9,13%				Erro médio quadrático (teste)	175,32
8-H	T reator (°C)	23,36%	38	05/06/2012	01/04/2014	Melhor rede	GRNN
	CQO reator (mg/L)	22,20%				% más previsões (treino - 30% toler.)	0,00%
	pH reator	18,74%				% más previsões (teste - 30% toler.)	37,50%
	AGV reator (mg/L)	18,51%				Erro médio quadrático (treino)	0,20
	TAL reator (mg/L)	17,20%				Erro médio quadrático (teste)	186,52
9-G-8	pH reator	37,32%	66	10/05/2012	06/05/2014	Melhor rede	GRNN
	T reator (°C)	33,51%				% más previsões (treino - 30% toler.)	0,00%
	TRH (dias)	21,05%				% más previsões (teste - 30% toler.)	38,46%
	LCQO in (kg/dia)	7,72%				Erro médio quadrático (treino)	26,46
	LMV in (kg/dia)	0,39%				Erro médio quadrático (teste)	201,57
9-G-1	pH reator	31,03%	41	03/04/2012	06/05/2014	Melhor rede	GRNN
	pH in	29,71%				% más previsões (treino - 30% toler.)	0,00%
	T reator (°C)	21,73%				% más previsões (teste - 30% toler.)	25,00%
	MV in (mg/L)	7,88%				Erro médio quadrático (treino)	15,95
	TRH (dias)	5,23%				Erro médio quadrático (teste)	136,95
	LMV in (kg/dia)	3,95%					
	AGV reator (mg/L)	0,32%					
	Q lamas (m³/dia)	0,16%					
10-A	pH reator	30,00%	44	20/03/2012	06/05/2014	Melhor rede	MLFN (5 nós)
	T reator (°C)	21,85%				% más previsões (treino - 30% toler.)	2,86%
	TRH (dias)	19,81%				% más previsões (teste - 30% toler.)	33,33%
	LMV in (kg/dia)	15,03%				Erro médio quadrático (treino)	23,04
	AGV reator (mg/L)	13,31%				Erro médio quadrático (teste)	214,82

Em relação à etapa de previsão, os resultados das redes neuronais exibidas na Tabela 5.5, podem ser consultados na Tabela 5.6.

Tabela 5.6 – Resultados da fase de previsão por rede neuronal (adaptado de Pinto, 2014)

Rede	Casos treino + teste	Casos previsão	RMS	R ²
8-F	42	11	105,6	0,45
8-H	38	11	123,5	0,12
9-G-8	66	12	124,3	0,05
9-G-1	41	11	182,8	0,07
10-A	44	11	180,8	0,01

Além dos resultados das cinco melhores redes neuronais geradas, considerou-se ainda importante recordar o resultado da rede “1-A” produzida a partir das variáveis “Q lamas” e “T reator” de frequência de amostragem diária. Esta rede possui um número de casos de treino e teste e de previsão bastante avultado, nomeadamente, 620 casos de treino/teste e 251 casos

de previsão. O erro médio quadrático registado (RMS) durante a fase de previsão foi de 163,19 e o coeficiente de determinação apresentou um valor de 0,422 (Pinto, 2014).

O desempenho da cada rede neuronal na fase de previsão é essencialmente avaliado pelo erro médio quadrático e o coeficiente de determinação (R^2). Quanto menor o erro médio quadrático, maior qualidade apresenta a rede neuronal, enquanto que o coeficiente de determinação deve ser o mais elevado possível, considerando-se que a previsão é muito boa, quando o valor do coeficiente é superior a 0,8 (Kasabov, 1998). Os resultados finais, tal como se indicou no trabalho anterior estão longe de serem bons, dado que a melhor rede neuronal na fase de previsão apresenta um erro médio quadrático acima do valor 100 e um coeficiente de determinação abaixo do valor 0,5.

As séries de dados produzidas no atual trabalho, reduzem o efeito de um dos fatores principais avançados como causa dos maus resultados verificados, nomeadamente, o facto de o número de dados disponível para a fase de treino e teste ser inferior ao desejado. Contudo, o objetivo principal da investigação para este primeiro caso de estudo passa por incorporar a variável condutividade nas séries de dados de previsão de valores de biogás produzido, pelo que nesta vertente do trabalho só haveria tempo para reproduzir algumas redes neuronais. A escolha mais correta considerou-se ser, voltar a reproduzir as cinco melhores redes neuronais do ano passado e a rede “1-A”. Esta última é uma rede interessante para analisar, dado que o número de casos utilizados na fase de treino e teste já era bastante elevado. No que diz respeito às redes neuronais “8-F”, “8-H”, “9-G-8”, “9-G-1” e “10-A”, estas voltaram-se a reproduzir com base nas novas séries de dados, incluindo a fase de previsão. Este processo ocorreu duas vezes para cada configuração, tendo-se produzido redes neuronais do tipo GRNN e MLFN com variações no número de nós selecionados, mas optando-se sempre por utilizar duas camadas ocultas.

O tempo de treino definido para cada rede foi de 30 minutos, sendo que em média cada grupo de redes neuronais demora pelo menos sete horas até a fase de treino e teste estar concluída. O resumo dos resultados globais da repetição da produção das redes neuronais pode ser consultado na Tabela 5.7. É importante realçar que se deu um novo nome às redes neuronais, de modo a evitar confusões entre as RNA criadas a partir das novas e antigas séries de dados.

Tabela 5.7 – Resultados globais da repetição da produção das cinco melhores redes neuronais de 2014

Rede	Variáveis entrada	Impacto	Nº casos	Data		Resultados	
8-F	AGV reator (mg/L)	25,84%	77	20/03/2012	25/02/2015	Melhor rede	MFLN (3 nós)
	TAL reator (mg/L)	24,45%				% más previsões (treino - 30% toler.)	11,29%
	pH reator	23,05%				% más previsões (teste - 30% toler.)	13,33%
	T reator (°C)	17,54%				Erro médio quadrático (treino)	85,20
	TRH (dias)	9,13%				Erro médio quadrático (teste)	85,09
8-H	T reator (°C)	34,59%	60	05/06/2012	29/10/2014	Melhor rede	MLFN (2 nós)
	CQO reator (mg/L)	23,98%				% más previsões (treino - 30% toler.)	22,92%
	TAL reator (mg/L)	20,24%				% más previsões (teste - 30% toler.)	33,33%
	AGV reator (mg/L)	11,32%				Erro médio quadrático (treino)	111,62
	pH reator	9,88%				Erro médio quadrático (teste)	133,35
9-G-8	T reator (°C)	23,12%	80	10/05/2012	29/10/2014	Melhor rede	MFLN (5 nós)
	TRH (dias)	23,08%				% más previsões (treino - 30% toler.)	3,12%
	pH reator	22,17%				% más previsões (teste - 30% toler.)	37,50%
	LCQO in (kg/dia)	16,32%				Erro médio quadrático (treino)	63,42
	LMV in (kg/dia)	15,30%				Erro médio quadrático (teste)	178,16
9-G-1	TRH (dias)	17,08%	72	03/04/2012	25/02/2014	Melhor rede	MFLN (3 nós)
	Q lamas (m³/dia)	12,71%				% más previsões (treino - 30% toler.)	12,07%
	pH reator	12,36%				% más previsões (teste - 30% toler.)	35,71%
	T reator (°C)	12,07%				Erro médio quadrático (treino)	56,33
	AGV reator (mg/L)	11,88%				Erro médio quadrático (teste)	137,26
	LMV in (kg/dia)	11,84%					
	pH in	11,33%					
	MV in (mg/L)	10,73%					
10-A	pH reator	40,45%	73	20/03/2012	25/02/2015	Melhor rede	GRNN
	T reator (°C)	34,16%				% más previsões (treino - 30% toler.)	15,52%
	LMV in (kg/dia)	14,58%				% más previsões (teste - 30% toler.)	20,00%
	AGV reator (mg/L)	7,41%				Erro médio quadrático (treino)	111,39
	TRH (dias)	3,38%				Erro médio quadrático (teste)	114,62

De modo geral, todas as novas séries de dados das redes neuronais (excetuando a “9-G-1 2015”) permitiram melhorar os resultados da fase de testes, o que indica em princípio uma maior capacidade de generalização e melhor apetência para a predição de valores. Esta melhoria tudo indica ser resultado do aumento do número de casos disponíveis para treino e testes, dado que foi esta a única alteração efetuada nas condições de criação das RNA. Os resultados da fase de previsão podem ser observados na Tabela 5.8. Os casos destacados para a fase de previsão datam de 18 de Março de 2015 a 26 de Agosto de 2015.

Tabela 5.8 – Resultados da fase de previsão da repetição da produção das cinco melhores redes neuronais de 2014

Rede	Casos treino + teste	Casos previsão	RMS	R ²
8-F 2015	77	10	223	0,48
8-H 2015	60	-	-	-
9-G-8 2015	80	-	-	-
9-G-1 2015	72	10	231	0,15
10-A 2015	73	10	179	0,10

Não foi possível executar a etapa de previsão nas redes neuronais “8-H 2015” e “9-G-8 2015”, visto estas terem sido treinadas com dados das variáveis “CQO reator” e “LCQO in”, sendo que à data dos casos de previsão utilizados, não são apresentados registos dos valores de CQO na ETAR do Seixal. Importa ainda realçar que o período temporal de onde saíram os casos de previsão se caracterizou por ser um período onde ocorreram alguns problemas operacionais, nomeadamente na afluência de lamas espessadas ao digestor anaeróbio. Estes problemas levaram a que durante este período os valores registados de algumas variáveis, como o “Q lamas” ou o “Q biogás” apresentassem valores anómalos e no global bastante dispersos, como pode ser observado no Anexo V. Este fator levou a que apenas se seleccionassem dez casos para integrar a previsão, de modo a tentar-se comprovar se o aumento do número de casos destacados para treino e teste permite ou não melhorar os resultados. A seleção de apenas alguns casos permitiu realizar a etapa de previsão em condições idênticas à realizada na investigação em 2014.

Os resultados registados da fase de previsão, no que diz respeito às três redes possíveis de testar, evidenciaram melhores resultados no método estatístico referente ao coeficiente de determinação e piores resultados no método estatístico relativo ao erro médio quadrático. Os resultados da fase de previsão não são claros, pelo que não se pode garantir que o aumento do número de dados tenha melhorado os resultados, porém é preciso ter em conta que o período de dados considerado para esta última etapa apresenta alguns problemas, apesar de se ter tentado minorá-los. Se entre Março e Agosto de 2015 não tivessem ocorridos problemas operacionais e os dados disponibilizados apresentassem valores mais consistentes, tudo leva a crer que os resultados da fase de previsão teriam seguido a tendência verificada nos resultados da fase de treino e teste, os quais melhoraram com o aumento do número de dados.

Por fim, reproduziu-se a rede “1-A 2015” a partir das novas séries de dados do conjunto de variáveis desta rede, mas ao contrário das redes anteriores, este processo apenas decorreu uma vez. A nova série de dados, criada a partir dos valores disponíveis de “Q lamas” e “T reator”, contém 942 casos de treino/teste. Foram destacados 106 casos para previsão, os quais se encontram entre as datas de 18 de Março de 2015 a 7 de Outubro de 2015. O coeficiente de determinação calculado foi de 0,00 e o erro médio quadrático de 214,11, ou seja, um mau resultado. A diferença deste resultado para o registado na investigação realizada em 2014 (resultado muito melhor), é difícil de explicar, tal como a diferença no coeficiente de correlação entre as variáveis “Q lamas” e “Q biogás” registadas entre trabalhos. Uma explicação plausível para esta diferença de resultados na etapa de previsão poderia ser o facto de, na investigação executada em 2014, os casos de previsão utilizados na RNA “1-A” terem também sido utilizados na fase de treino e teste, contudo não se tem conhecimento se assim foi. Este facto faz toda a diferença, se no presente os casos selecionados para a etapa de treino e teste da rede “1-A 2015” tivessem sido selecionados para predição de valores, o resultado do coeficiente de determinação teria sido de 0,58 e o erro médio quadrático de 122,00, o que representaria uma melhoria.

Este dado levou a que considerasse importante apresentar os resultados da fase de previsão das redes “8-F 2015”, “8-H 2015”, “9-G-8 2015”, “9-G-1 2015” e “10-A 2015”, quando os dados destacados para treino e teste também o são para a etapa de predição de valores. Os resultados podem ser conferidos na Tabela 5.9.

Tabela 5.9 – Resultados quando destacados casos para previsão, anteriormente utilizados nas fases de treino e teste de uma rede neuronal

Rede	Casos treino + teste	Casos previsão	RMS	R ²
8-F 2015	77	32	94,5	0,83
8-H 2015	60	18	116,4	0,58
9-G-8 2015	80	18	85,6	0,77
9-G-1 2015	72	32	87,4	0,78
10-A 2015	73	32	111,9	0,64

Em suma, após a adição de novos dados às séries de dados, depois de geradas novas redes neuronais com base nas novas séries, e efetuada a previsão de valores em situações diferentes, é possível retirar as seguintes ilações:

- Muito mais importante que a quantidade de dados a considerar na produção de RNA é a qualidade desses mesmos dados;
- O facto de que as séries de dados antigas possuíam poucos dados é inegável, mesmo no presente este aspeto continua a ser uma realidade das novas séries de dados afetas às variáveis de frequência de amostragem semanal, as quais alimentaram as RNA que geraram melhores resultados. É credível que com o aumento de dados, os resultados melhorem, tal como se verificou na previsão das redes neuronais, quando os casos destacados para treino e teste também foram usados na fase de previsão;
- Contudo, a forma mais correta de avaliar a capacidade de generalização de uma RNA é destacar casos para previsão que não foram utilizados nas etapas de treino e teste de uma rede neuronal. Ao utilizar-se os mesmos casos nas três fases, estamos a inflacionar os resultados finais, dado que a rede durante o treino conseguiu assimilar grande parte dos padrões existentes nas séries de dados. Dar a uma RNA a tarefa de prever padrões para os quais foi treinada é naturalmente mais fácil e tudo indica que quanto maior o número de dados disponível, desde que a variabilidade não aumente em uma proporção fora do normal, melhor serão os resultados;
- O mesmo pode não acontecer nas situações em que os casos destacados para previsão são novos, nestas situações o problema é muito mais complexo, dado que o principal fator que irá influenciar os resultados é a própria variabilidade do processo em estudo que na digestão anaeróbia é bastante elevada. No atual caso, os dados adicionados não tiveram um grande impacto na variabilidade das séries de dados, pelo que o aumento de dados deveria ser benéfico. Todavia os casos destacados para previsão foram retirados de um

período temporal em que ocorreram alguns problemas operacionais, pelo que este fator pode ter afetado os resultados.

5.1.3 Análise da Influência de Intrusões Salinas no Processo de Digestão Anaeróbia

Tendo em vista a análise da influência das intrusões salinas (detetadas no efluente que chega à ETAR do Seixal) no funcionamento do processo de digestão anaeróbia, o objetivo principal a concretizar na investigação, era incorporar a condutividade como uma variável a ter em conta na previsão de valores de biogás produzidos. Os dados de condutividade média diária disponíveis até ao momento eram poucos e não eram suficientes para serem associados a outras variáveis, de modo a gerar RNA capazes de produzir resultados de carácter significativo. Para o objetivo principal ser cumprido era preciso aumentar o número de dados, algo que se considerava ser legítimo caso fosse possível desenvolver uma rede neuronal capaz de prever valores de condutividade. A rede neuronal desenvolvida poderia estimar os valores de condutividade existentes no interior do digestor durante o último ano, o que permitiria a obtenção de dados suficientes para esta variável ser adicionada ao conjunto até agora usado na predição de valores de produção de biogás.

5.1.3.1 Recolha e Preparação de dados

O primeiro passo de desenvolvimento de uma rede neuronal com capacidade de estimar valores de condutividade, foi a recolha de dados de condutividade e suas variáveis afetas. A construção das séries de dados aconteceu de modo intermitente, contudo, é apresentado de seguida um sumário dos dados totais disponibilizados.

Relativamente aos dados de condutividade, foram disponibilizados 3265 valores de modo faseado, ou seja, um número elevado, mas que apenas representa medições de 36 dias, dado que o intervalo temporal entre o registo de valores é de apenas 15 minutos. O número de dias em potencial é de 46, mas só foram considerados como dias válidos os que apresentavam no mínimo 12 horas de valores registados. O número de dias considerado levanta o problema de articulação com outras variáveis para a previsão de valores de biogás produzido. As RNA com melhor capacidade de predição de valores de biogás produzido, ao serem geradas a partir de variáveis de valores diários de frequência de amostragem semanal ou mesmo inferior a uma vez por semana, impossibilita a adição da variável condutividade ao conjunto. Em caso de adição, os casos disponíveis para treinamento de rede seriam apenas quatro, não sendo possível a partir destes criar uma rede neuronal de resultados significativos. A média dos valores disponíveis de condutividade é de cerca 6,85 mS/cm, o que indica claramente a existência de intrusões salinas. É importante relembrar que se considera a ocorrência de intrusões salinas no efluente da ETAR do Seixal, a partir de valores de condutividade registados superiores a 2 mS/cm (Serrano, 2014).

As variáveis explicativas dos valores de condutividade verificados na ETAR do Seixal selecionadas para fazer parte das séries de dados, que serviriam de base ao desenvolvimento das RNA para predição de valores de condutividade, foram a altura de maré e o caudal instantâneo. No que diz respeito aos valores de altura de maré, utilizaram-se as equações 3.4 e 3.5 para a concretização de várias curvas de maré, dado ser necessário conhecer a altura de maré no instante a que foi registado cada valor de condutividade. A execução das curvas de maré exigiu a recolha de 82 valores de preia-mar e 79 valores de baixa-mar. A inconformidade apurada entre o número de valores de preia-mar e baixa-mar recolhidos, advém das interrupções temporais que por vezes existiam no registo de dados de condutividade, levando a que nem sempre fosse necessário recolher um valor de cada. O critério das 12 horas que definiu o número de dias válidos em relação à série de dados de condutividade, tem em conta o facto de o intervalo entre uma preia-mar e a seguinte ser de aproximadamente 12 horas e 25 minutos (Instituto Hidrográfico, 2015). Caso os valores de condutividade fossem usados para previsão de valores de biogás, se se considerasse a média de valores de dias com medições de menos de 12 horas, os padrões existentes nos dados poderiam ser adulterados, o que prejudicaria a aprendizagem das redes neuronais. Porém, para a predição de valores de condutividade, todos os valores disponibilizados foram considerados para as fases de treino e teste das RNA, incluindo os dos dias não válidos.

A variável “Q instantâneo” permite conhecer o valor de caudal associado no momento do registo de cada valor de condutividade. O número de dados disponíveis de caudal instantâneo é inferior ao número de dados de condutividade e altura de maré, pelo que nas RNA em que a série de dados desta variável é utilizada, não é possível utilizar as restantes séries de dados na sua capacidade máxima. Apesar da limitação imposta pelo número de dados de caudal instantâneo disponíveis, não é provável que esta afete os resultados gerados, dado o número de casos disponíveis continuar a ser bastante elevado. Não foi possível correlacionar mais variáveis aos dados de condutividade, visto não se ter conhecimento de outras variáveis afetas e as quais possuam valores registados de forma contínua em um curto espaço de tempo.

Na Figura 5.5 é possível visualizar a relação entre os valores de condutividade e os valores da altura de maré.

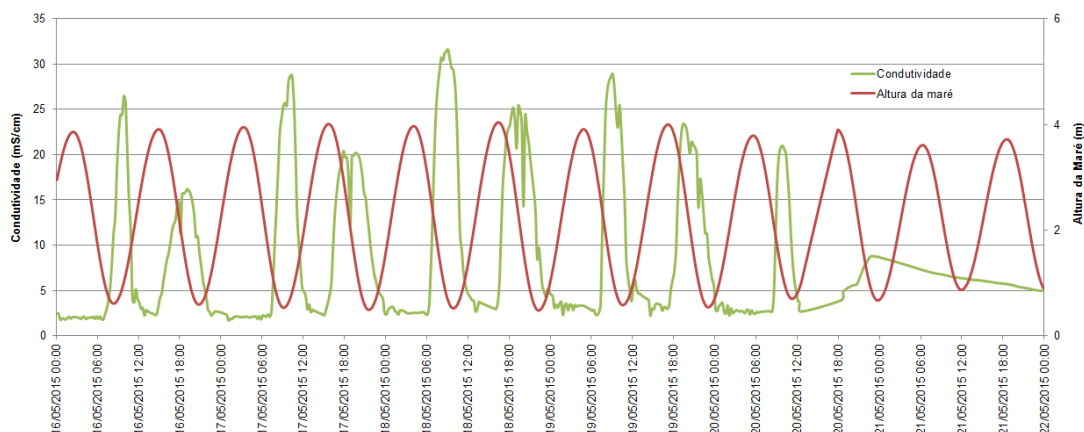


Figura 5.5 – Dados de condutividade e da altura de maré de 16 de Maio a 21 de Maio de 2015

Ao observarmos a Figura 5.5, é possível constatar que existe uma clara influência da altura de maré nos valores de condutividade registados. Os valores de condutividade são mais elevados quando a altura de maré é mais elevada, enquanto os valores de condutividade mais baixos são registados quando a altura de maré é mais reduzida. Contudo, é possível verificar que os picos dos valores de condutividade acontecem sistematicamente entre três a seis horas depois dos picos de altura de maré, o que indica a existência de um atraso horário do efeito da altura de maré nos valores de condutividade.

Na Figura 5.6 é exequível a observação da relação entre os valores de condutividade e os valores de caudal instantâneo. Todavia os valores de caudal instantâneo são apresentados em formato de caudal médio horário, para que a observação do gráfico seja acessível. Em ambos os gráficos, o período temporal diz respeito ao período temporal máximo para o qual existem dados em contínuo disponíveis (de 16 de Maio de 2015 a 21 de Maio de 2015).

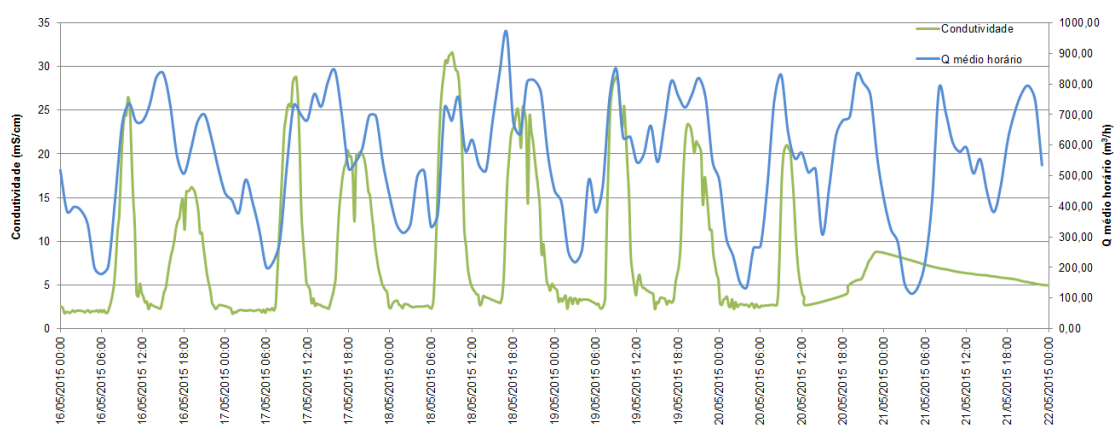


Figura 5.6 - Dados de condutividade e de caudal médio horário de 16 de Maio a 21 de Maio de 2015

A análise da Figura 5.6 sugere que existe uma relação entre a condutividade registada e o caudal médio horário registado. Os períodos temporais de valores elevados de condutividade

condizem de forma mais-ou-menos comum com os períodos temporais em que os valores de caudal também são maiores, embora a correlação não pareça ser tão forte como a existente na Figura 5.5.

O coeficiente de correlação de Pearson foi calculado de forma a possibilitar uma análise mais correta das correlações entre série de dados. O valor de correlação calculado entre a série de dados totais de condutividade e a série de dados totais da altura de maré apresentou um valor 0,43. O cálculo do coeficiente entre a série de dados totais de condutividade e a série de dados totais de caudal instantâneo revelou um valor de 0,25. Apesar de existir uma correlação moderada entre os dados de condutividade e os dados da altura de maré, pensava-se à partida, e com base na Figura 5.5, que a correlação entre estas duas variáveis seria mais forte. Este valor de correlação não muito forte indica que provavelmente existirão outras variáveis desconhecidas, que poderão exercer uma pequena influência sobre os valores de condutividade registados, levando a que por vezes existam algumas flutuações nos valores registados de condutividade.

Por último no que diz respeito aos dados totais disponibilizados, quis-se perceber se existe alguma influência da sazonalidade nos valores registados. Esta análise, no entanto não pode ser encarada como definitiva, dado que dos 46 dias de dados considerados, apenas seis dizem respeito ao período de Inverno, enquanto os restantes se encontram todos entre os dias 15 de Maio e 28 de Julho. Na Tabela 5.10 mostram-se os valores médios de cada variável nos dois períodos temporais referidos.

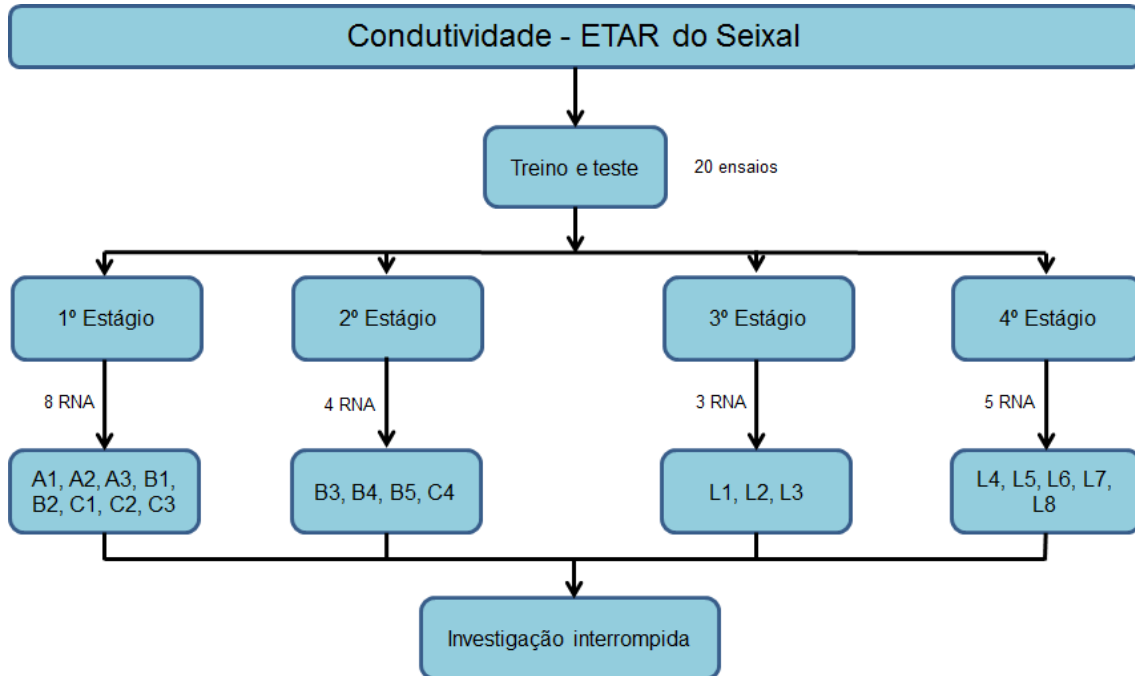
Tabela 5.10 – Valores médios de condutividade, altura de maré e caudal instantâneo, por período temporal

Período	Nº Dados	Média		
		Condutividade (mS/cm)	Altura de maré (m)	Caudal instantâneo (m ³ /s)
Inverno	493	8,52	2,04	570,37
Primavera/Verão	2772	6,55	2,08	446,66

Nos dados analisados, nota-se claramente um aumento nos valores de condutividade e nos valores de caudal instantâneo na época de Inverno. As médias aferidas indicam que deverá existir variações devido à sazonalidade, no entanto, os dados disponíveis, referente ao período de Inverno são escassos em comparação aos de final da Primavera e de Verão, não se podendo tirar conclusões definitivas. Detetada esta pequena diferença entre ordem de valores, pensou-se em retirar os dados de Inverno e testar redes neuronais somente com os restantes dados. No entanto, ao se retirar os dados da época invernal, o coeficiente de correlação entre a série de dados de condutividade e a série de dados entre os dados da altura de maré baixou, pelo que se decidiu não fazer distinção pela época sazonal a que cada dado pertence.

5.1.3.2 Etapa de Treino e Teste

A etapa de treino e teste realizada é um pouco complexa, pelo que antes de ser apresentada detalhadamente, é possível observar em esquema na Figura 5.7 o resumo desta fase da investigação.



Legenda:
A – RNA de familiarização
B – RNA de série de dados contínuas, sem seleção de dados
C – RNA de série de dados com desfasamento, sem seleção de dados
L – RNA com seleção de dados

Figura 5.7 – Metodologia aplicada na etapa de treino e teste das redes neuronais criadas com o intuito de prever valores de condutividade na ETAR do Seixal

Cada RNA desenvolvida foi sujeita a uma fase de treinamento e a uma fase de testes (as fases ocorreram em simultâneo), tendo em consideração o objetivo de previsão dos valores de condutividade. Foram criadas várias redes neuronais, porém estas não foram desenvolvidas tendo por base sempre o mesmo conjunto de dados, visto os valores de condutividade não terem sido disponibilizados todos ao mesmo tempo. Ainda assim, a chegada dos dados de modo intermitente não afetou a produção das RNA nem os resultados gerados. Durante este processo foram treinadas e testadas vinte redes neuronais diferentes. Apesar de existir apenas duas variáveis independentes e uma dependente, foram testadas RNA em diferentes condições, de modo a se tentar produzir o melhor resultado possível. A fase de treino e testes pode ser dividida em quatro estágios.

O primeiro estágio diz respeito às RNA produzidas com a primeira versão da série de dados criada. Este conjunto de dados englobava apenas registos de cinco dias, porém o número de casos disponíveis para a produção das redes neuronais já era bastante elevado. O coeficiente

de correlação de Pearson, neste conjunto de dados, entre os valores de condutividade e os de altura de maré é de 0,68, enquanto entre os valores de condutividade e os de Q instantâneo é 0,21. A partir deste conjunto de dados foram criadas oito redes neuronais: “A1”; “A2”; “A3”; “B1”; “B2”; “C1”; “C2”; “C3”. As redes da classe A, ao contrário das da classe B e C, foram criadas mais com um intuito de familiarização do que com o intuito de produzir resultados significativos. As redes da classe B foram geradas tendo por base uma série de dados contínua sem qualquer modificação ou seleção de dados. A única diferença da rede “B1” para a rede “B2”, consiste no facto de na “B2” apenas se ter usado uma variável independente, de modo a se constatar se a ausência da série de dados de caudal instantâneo melhorava ou piorava os resultados produzidos pela “B1”. Os resultados produzidos por estas redes neuronais podem ser consultados na Tabela 5.11.

Tabela 5.11 – Resultados gerados pelas redes neuronais de classe B

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
B1	Altura maré (m)	68,70%	398	21/01/2015	26/01/2015	Melhor rede	MLFN (9 nós)
	Q instantâneo (m³/s)	32,30%				% más previsões (treino - 30% toler.)	52,83%
						% más previsões (teste - 30% toler.)	50,00%
						Erro médio quadrático (treino)	3,63
						Erro médio quadrático (teste)	5,31
B2	Altura maré (m)	100,00%	493	21/01/2015	26/01/2015	Melhor rede	MFLN (2 nós)
						% más previsões (treino - 30% toler.)	61,67%
						% más previsões (teste - 30% toler.)	58,59%
						Erro médio quadrático (treino)	5,27
						Erro médio quadrático (teste)	5,15

Após observação da Tabela 5.11, é perceptível que os resultados não foram bons e que a ausência da série de dados da variável Q instantâneo produziu piores resultados. Tendo em conta os resultados da classe B neste estágio e o facto de a altura de maré ter um efeito tardio nos valores de condutividade registados, decidiu-se criar redes neuronais com desfaseamento horário entre os dados das variáveis. A rede “C1” apresenta um desfaseamento de três horas da variável condutividade face às restantes, enquanto a rede “C2” exhibe o mesmo desfaseamento horário, mas neste caso, além da variável condutividade, também a variável “Q instantâneo” está desfasada em relação à altura de maré. A concepção da rede “C3” está baseada em um desfaseamento de quatro horas dos valores de condutividade face aos valores das outras duas variáveis existentes. Os resultados das redes de classe C geradas neste primeiro estágio, podem ser consultados na Tabela 5.12.

Tabela 5.12 – Resultados gerados pelas redes neuronais de classe C

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
C1	Altura maré (m)	57,00%	388	21/01/2015	26/01/2015	Melhor rede	MFLN (7 nós)
	Q instantâneo (m ³ /s)	43,00%				% más previsões (treino - 30% toler.)	60,00%
						% más previsões (teste - 30% toler.)	71,49%
						Erro médio quadrático (treino)	5,35
						Erro médio quadrático (teste)	6,29
C2	Altura maré (m)	77,75%	388	21/01/2015	26/01/2015	Melhor rede	MFLN (3 nós)
	Q instantâneo (m ³ /s)	22,25%				% más previsões (treino - 30% toler.)	68,71%
						% más previsões (teste - 30% toler.)	74,36%
						Erro médio quadrático (treino)	5,86
						Erro médio quadrático (teste)	6,98
C3	Altura maré (m)	72,33%	383	21/01/2015	26/01/2015	Melhor rede	GRNN
	Q instantâneo (m ³ /s)	27,67%				% más previsões (treino - 30% toler.)	56,86%
						% más previsões (teste - 30% toler.)	63,64%
						Erro médio quadrático (treino)	4,96
						Erro médio quadrático (teste)	4,46

Os resultados da Tabela 5.12 indicam que o desfasamento horário das variáveis não foi a solução adequada, tendo em vista a melhoria dos resultados. No geral os resultados foram piores em relação às redes de classe B.

O segundo estágio da criação das RNA começou após a chegada de novos dados afetos às variáveis. Este estágio engloba séries de dados de 37 dias, o que se aproxima bastante do máximo de dados disponibilizados no total. O cálculo do coeficiente de Pearson para estas séries de dados evidenciou um valor de 0,45 entre a série condutividade e a série altura de maré e um valor de 0,25 entre a série condutividade e o caudal instantâneo. Nesta etapa, foram geradas as seguintes redes neuronais: “B3”; “B4”; “B5”; “C4”. As redes de classe B foram criadas de forma a continuar a estudar-se a performance das séries de dados contínuas. A rede “B3” é idêntica à rede “B1”, sendo que no caso da rede “B3”, o número de dados que possui é muito mais elevado, por isso esta rede serviu para comprovar se o aumento de dados poderia ou não produzir melhores resultados. Na rede “B4” decidiu-se eliminar os valores de caudal instantâneo abaixo dos 100 m³/s, dado que os mesmos se apresentam bastante abaixo da média, o que afeta a consistência das séries de dados. A rede “B5” é idêntica à rede “B2”, apesar de esta última ter tido piores resultados do que a rede “B1”, decidiu-se testar novamente uma configuração sem a variável “Q instantâneo”, procurando-se examinar se o aumento do número de casos melhoraria os resultados desta configuração. Os resultados das RNA de classe B geradas durante este segundo estágio podem ser consultados na Tabela 5.13.

Tabela 5.13 – Resultados das redes neuronais de classe B, referentes ao segundo estágio

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
B3	Altura maré (m)	81,70%	1690	15/05/2014	26/01/2015	Melhor rede	MFLN (6 nós)
	Q instantâneo (m³/s)	18,30%				% más previsões (treino - 30% toler.)	74,56%
						% más previsões (teste - 30% toler.)	73,08%
						Erro médio quadrático (treino)	4,77
						Erro médio quadrático (teste)	4,72
B4	Altura maré (m)	78,81%	1389	15/05/2014	26/01/2015	Melhor rede	GRNN
	Q instantâneo (m³/s)	21,19%				% más previsões (treino - 30% toler.)	71,83%
						% más previsões (teste - 30% toler.)	72,66%
						Erro médio quadrático (treino)	4,87
						Erro médio quadrático (teste)	5,05
B5	Altura maré (m)	100,00%	2644	15/05/2014	26/01/2015	Melhor rede	MFLN (10 nós)
						% más previsões (treino - 30% toler.)	74,89%
						% más previsões (teste - 30% toler.)	72,40%
						Erro médio quadrático (treino)	5,29
						Erro médio quadrático (teste)	5,68

Com o aumento de casos, os resultados não melhoraram como se pode observar na Tabela 5.13. A remoção dos valores de caudal instantâneo abaixo dos 100 m³/s também não fez grande diferença, tal como a remoção total da série de dados de caudal instantâneo também não proporcionou melhores resultados.

Neste estágio, foi ainda testada a rede “C4”, a qual detém a série de dados de condutividade desfasada seis horas das remanescentes séries de dados. Na análise da Tabela 5.14, a qual apresenta o resultado gerado pela rede “C4”, esta também não apresentou melhores resultados face às redes de classe B. No geral todas as redes de desfasamento criadas até agora apresentaram resultados piores, quando comparadas às redes geradas nas mesmas condições, mas sem desfasamento de valores.

Tabela 5.14 – Resultados gerados pela rede neuronal “C4”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
C4	Altura maré (m)	82,53%	1368	15/05/2014	26/01/2015	Melhor rede	GRNN
	Q instantâneo (m³/s)	17,47%				% más previsões (treino - 30% toler.)	78,70%
						% más previsões (teste - 30% toler.)	75,18%
						Erro médio quadrático (treino)	4,91
						Erro médio quadrático (teste)	5,79

Numa análise global das redes geradas nos dois estágios até agora apresentados, o problema não está nas configurações utilizadas, mas sim no facto de existir uma grande variância na ordem de valores em cada série de dados, o que complica a aprendizagem das RNA. O desvio padrão das séries de dados do segundo estágio é de 6,75 para a série de dados de condutividade, 1,03 para a série de dados da altura de maré e atinge um valor na ordem dos 300 para a série de dados de caudal instantâneo. Os resultados do cálculo do desvio padrão indicam a grande variedade existente na gama de valores de cada variável (com exceção da altura de maré). Tendo em conta que as RNA estão a ser treinadas com base em dados de três

variáveis, esta variabilidade de valores complica bastante a aprendizagem de qualquer padrão que possa existir nas séries de dados.

A principal preocupação na concepção das RNA no terceiro estágio, foi diminuir a discrepância de valores, ou seja, eliminar os valores que mais fogem ao padrão comum das séries de dados. A solução passou por executar uma seleção de dados, a qual teve por base os seguintes critérios:

- Removeram-se os valores de caudal instantâneo abaixo de 150 m³/s e acima de 1160 m³/s. Os valores removidos não se apresentam dentro da normalidade do funcionamento da ETAR;
- A partir de uma função simples, com base na divisão dos valores de condutividade pelos valores da altura de maré, verificou-se que regra geral, os valores mais altos de condutividade registados correspondem a baixos valores de altura de maré e os valores mais baixos a elevados valores de altura de maré (isto devido ao atraso do efeito da altura de maré nos valores de condutividade observados). Os resultados gerados pelo cálculo da função utilizada permitiram eliminar os casos que mais fugiam ao padrão existente;
- Com base nos dois critérios anteriores, escolheu-se eliminar dados até se remover aproximadamente 30% dos dados totais.

Depois de executada a eliminação de praticamente 30% dos dados criou-se a rede neuronal “L1”. Os resultados da rede “L1” podem ser consultados na Tabela 5.15. A série de dados que alimentou a produção desta rede apresenta as seguintes características:

- O cálculo do coeficiente de correlação de Pearson apresenta um resultado de 0,60 entre a série de dados de condutividade a série de dados da altura de maré e um resultado de 0,25 entre a série de dados de condutividade e a série de dados de caudal instantâneo. Estes resultados representam um aumento significativo do coeficiente de correlação entre a série de dados de condutividade e a série de dados da altura de maré face ao registado anteriormente;
- O desvio padrão registado na série de valores de condutividade é 6,70; na série de valores de altura de maré é de 1,03 e na série de dados de caudal instantâneo é aproximadamente de 240. Existe sobretudo uma redução no desvio padrão dos dados de caudal instantâneo.

Tabela 5.15 – Resultados gerados pela rede neuronal “L1”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
L1	Altura maré (m)	81,37%	1189	15/05/2014	26/01/2015	Melhor rede	MFLN (6 nós)
	Q instantâneo (m ³ /s)	18,63%				% más previsões (treino - 30% toler.)	63,51%
						% más previsões (teste - 30% toler.)	66,38%
						Erro médio quadrático (treino)	3,92
						Erro médio quadrático (teste)	4,26

Os resultados da rede “L1” denotam uma ligeira melhoria em comparação aos resultados produzidos pelas redes de segundo estágio, no entanto, são inferiores aos da rede “B1” e estão longe de poderem ser considerados bons.

Neste terceiro estágio quis-se ainda entender se o período horário tinha alguma influência nos resultados. Dado que os valores de condutividade e altura de maré são mais ou menos cíclicos, a única variável que poderia influenciar esta temática seria o “Q instantâneo”. Executou-se a média horária dos valores de caudal instantâneo disponibilizados, os quais geraram o gráfico representado na Figura 5.8.

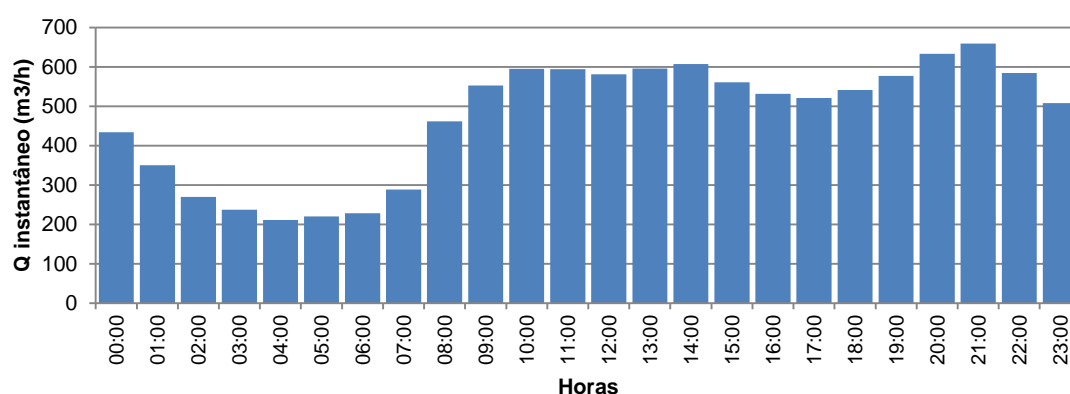


Figura 5.8 – Média horária dos valores de “Q instantâneo”

A análise efetuada ao gráfico da Figura 5.8 permitiu definir dois períodos, um das 00h às 8h59min e outro das 09h às 23h59min. O passo seguinte foi a produção de duas RNA com base nos dois períodos horários definidos, as quais se designaram por “L2” e “L3”. A rede neuronal “L2” tem por base os dados referentes ao período horário das 0h às 8h59min, enquanto a rede neuronal “L3” tem por base os dados das 09h às 23h59min. Em ambos os casos, foram aplicados os mesmos critérios que se utilizaram na preparação das séries de dados que alimentaram a rede “L1”. Os resultados gerados por estas duas redes podem ser observados na Tabela 5.16.

Tabela 5.16 – Resultados gerados pelas redes neuronais “L2” e “L3”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
L2	Altura maré (m)	58,92%	250	15/05/2014	26/01/2015	Melhor rede	GRNN
	Q instantâneo (m³/s)	41,08%				% más previsões (treino - 30% toler.)	54,00%
						% más previsões (teste - 30% toler.)	52,00%
						Erro médio quadrático (treino)	2,09
						Erro médio quadrático (teste)	2,52
L3	Altura maré (m)	83,49%	910	15/05/2014	26/01/2015	Melhor rede	GRNN
	Q instantâneo (m³/s)	16,51%				% más previsões (treino - 30% toler.)	47,25%
						% más previsões (teste - 30% toler.)	53,85%
						Erro médio quadrático (treino)	3,98
						Erro médio quadrático (teste)	4,52

Os resultados gerados pelas redes neuronais com separação horária de dados geraram melhores resultados em relação aos resultados da rede neuronal sem separação horária de dados. Os resultados são substancialmente melhores em separado, pois ao, dividir-se as séries de dados em dois grupos por médias horárias, está-se também a diminuir a variabilidade de valores existentes, o que facilita o processo de aprendizagem das RNA.

No último estágio, foram geradas cinco RNA, “L4”, “L5”, “L6”, “L7” e “L8”, sendo que a base de dados de cada uma possui todos os dados disponibilizados para tratamento de informação. A rede neuronal “L4” é semelhante à rede neuronal “L1”, dado que foram aplicados os mesmos critérios na seleção de casos válidos para a fase de treino e teste, mas no caso da rede “L4” esta possui um maior número de casos. Esta rede neuronal foi produzida com o intuito de comprovar se o aumento do número de casos, após aplicação de critérios de seleção de dados, piorava ou melhorava os resultados. Nas opções de concepção da rede neuronal “L5”, decidiu-se prolongar o tipo de redes concebidas até às redes MFLN de 30 nós, pois até agora tinham-se criado apenas redes MFLN até 15 nós. Na configuração da rede neuronal “L6”, elevou-se o tempo de treino e teste de cada rede até 1h. As redes produzidas anteriormente foram treinadas e testadas durante 30min. Estas alterações de concepção serviram para perceber se era possível produzir melhores resultados, com base apenas em alterações nas condições de treino. As redes neuronais “L7” e “L8” foram geradas, tendo por base as mesmas condições de criação das redes neuronais “L2” e “L3”, respetivamente, apenas detêm séries de dados maiores. A produção destas redes foi com o intuito de saber se o aumento de dados piorava ou melhorava os resultados, em casos em que existe separação horária nas séries de dados. A compilação dos resultados originados pelas redes neuronais criadas neste último estágio pode ser consultada na Tabela 5.17.

Tabela 5.17 – Resultados das redes neuronais geradas no quarto estágio

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
L4	Altura maré (m)	84,98%	1719	15/05/2014	22/05/2015	Melhor rede	MFLN (3 nós)
	Q instantâneo (m ³ /s)	15,02%				% más previsões (treino - 30% toler.)	68,80%
						% más previsões (teste - 30% toler.)	66,57%
						Erro médio quadrático (treino)	4,90
						Erro médio quadrático (teste)	4,79
L5	Altura maré (m)	83,38%	1719	15/05/2014	22/05/2015	Melhor rede	MFLN (21 nós)
	Q instantâneo (m ³ /s)	16,62%				% más previsões (treino - 30% toler.)	63,71%
						% más previsões (teste - 30% toler.)	67,15%
						Erro médio quadrático (treino)	4,61
						Erro médio quadrático (teste)	5,27
L6	Altura maré (m)	85,38%	1719	15/05/2014	22/05/2015	Melhor rede	MFLN (3 nós)
	Q instantâneo (m ³ /s)	14,62%				% más previsões (treino - 30% toler.)	68,95%
						% más previsões (teste - 30% toler.)	66,28%
						Erro médio quadrático (treino)	4,85
						Erro médio quadrático (teste)	4,97
L7	Altura maré (m)	63,13%	435	15/05/2014	22/05/2015	Melhor rede	GRNN
	Q instantâneo (m ³ /s)	36,87%				% más previsões (treino - 30% toler.)	61,49%
						% más previsões (teste - 30% toler.)	67,82%
						Erro médio quadrático (treino)	3,98
						Erro médio quadrático (teste)	4,87
L8	Altura maré (m)	93,08%	1255	15/05/2014	22/05/2015	Melhor rede	GRNN
	Q instantâneo (m ³ /s)	6,92%				% más previsões (treino - 30% toler.)	66,73%
						% más previsões (teste - 30% toler.)	65,34%
						Erro médio quadrático (treino)	4,87
						Erro médio quadrático (teste)	5,34

A rede neuronal “L4” teve resultados ligeiramente piores que a rede “L1”, registando-se mais uma vez resultados piores após o aumento do número de dados, mesmo utilizando, neste caso, critérios de eliminação de dados. Nas redes neuronais “L5” e “L6”, apesar de se ter aumentado, respetivamente, o número de tipos de redes testadas e o tempo de treinamento, não se verificaram melhorias nos resultados. Os resultados das RNA “L5” e “L6” indicam que provavelmente não valerá a pena aumentar o número de tipos de redes testadas e o tempo de treino, quando os resultados à partida estão longe de se poderem considerar bons. Esta opção por prolongar a fase de treino e teste, retira algum tempo útil de trabalho, pelo que, o mais indicado será apenas utilizar esta opção em casos de estudo onde se registam bons resultados, valendo a pena aprofundar os testes realizados com as melhores RNA. O aumento do número de dados das séries que alimentaram a produção das redes neuronais “L7” e “L8”, conduziu a que estas gerassem piores resultados do que as redes “L2” e “L3”.

O quarto estágio marcou o culminar da criação de RNA com o objetivo de prever valores de condutividade. Tendo por base os resultados gerados nos diferentes estágios, registam-se as seguintes conclusões:

- Os resultados gerados foram maus, pelo que não foi possível seleccionar nenhuma rede para a fase de predição de valores de condutividade;
- O facto de se não ter conseguido produzir uma rede neuronal com boa capacidade de previsão de valores de condutividade a partir das variáveis utilizadas, indica que esta é uma tarefa difícil de concretização num futuro próximo;

- Os resultados gerados pelas RNA com séries de dados desfasadas umas das outras, produziram piores resultados que as RNA de séries de dados sem desfasamento. No entanto, se no futuro for possível aumentar o número de variáveis afetas à condutividade, as configurações de séries de dados com desfasamento devem continuar a ser testadas;
- Os testes realizados evidenciaram que à medida que o número de casos aumentava, os resultados pioravam. Esta ocorrência verificada, provavelmente tem por base o facto de só se terem usado duas variáveis independentes no treino das diferentes RNA e o facto de estas séries de dados não conseguirem explicar totalmente os valores de condutividade registados. Qualquer incremento do número de dados, acabava por gerar um aumento na variabilidade dos valores registados, o que neste caso afetou claramente a aprendizagem das RNA. Se estivessem disponíveis mais variáveis, existe a hipótese de a ampliação do número de dados ser benéfica. Contudo em situações em que existem poucas variáveis disponíveis e as séries de dados são pouco estáveis, pensa-se que o aumento de dados poderá frequentemente piorar os resultados. Porém não faz sentido gerar RNA com base em poucos dados, visto a generalização que se iria produzir ser extremamente pouco significativa e não se poderia aplicar na prática. Perante circunstâncias em que existem poucas variáveis disponíveis e as séries de dados são pouco estáveis, pensa-se ser muito difícil criar uma rede neuronal capaz de conceber uma boa generalização e de aplicação real.

O objetivo de aumentar o número de dados diários de condutividade não foi atingido, dado não ter sido possível criar uma rede neuronal com boa capacidade de previsão de valores de condutividade. O facto de na ETAR ainda não ter sido possível medir a condutividade no caudal de lamas afluente à digestão anaeróbia de forma contínua, também não permitiu avaliar o impacto deste parâmetro na qualidade das lamas. Estes dois fatores tornaram impossível concretizar o objetivo de utilizar valores de condutividade como uma variável a ter em conta, na previsão de valores de produção de biogás durante o processo de digestão anaeróbia. A impossibilidade verificada não permitiu fazer mais progressos na análise dos efeitos das intrusões salinas no funcionamento e eficiência do digestor anaeróbio, nem continuar a desenvolver a investigação concretizada na primeira vertente deste caso de estudo.

5.2 Caso de Estudo – ETAR de Sesimbra

O principal objetivo do estudo realizado na ETAR de Sesimbra, tal como se sucedeu na ETAR do Seixal, passa por utilizar as potencialidades de modelação de processos das RNA no funcionamento da etapa de digestão anaeróbia desta instalação.

5.2.1 Recolha e Preparação de Dados

O primeiro passo dado no desenvolvimento da investigação na ETAR de Sesimbra foi consultar as séries de registo disponibilizadas, de modo a se tomar conhecimento das variáveis

disponíveis e afetas ao processo de digestão anaeróbia. A informação disponível neste caso de estudo era semelhante à informação disponibilizada nas séries de registo da ETAR do Seixal, as principais diferenças encontradas dizem respeito ao número de variáveis disponíveis e à profundidade temporal dos registos. Na ETAR do Seixal existiam 13 variáveis de medição ou análise laboratorial afetas ao processo de digestão anaeróbia, enquanto na ETAR de Sesimbra existem apenas 11. Nas séries de registo da ETAR de Sesimbra não existem dados das variáveis “CQO in” e “CQO reator”, pelo que também não foi possível calcular valores para as variáveis “LCQO in” e “LCQO out”. Em relação às restantes séries de dados, além das que foram usadas na ETAR do Seixal, no caso atual utilizou-se também a variável “Q efluente bruto”, de forma a verificar se a adição desta variável permitiria obter melhores resultados nas RNA produzidas a partir de séries de dados de frequência de amostragem diária. Os registos consultados, neste caso, datam de 1 de Janeiro de 2013 a 3 de Agosto de 2015, um período inferior ao que serviu de base à investigação no primeiro caso de estudo. Na Tabela 5.18 é possível constatar algumas informações das variáveis consideradas para integrar o estudo. Os locais de onde foram retiradas as amostras para laboratório e medidos alguns parâmetros são apresentados no Anexo VI.

Tabela 5.18 – Variáveis consideradas para integrar o estudo

Variável	Unidade	Frequência de amostragem	Número de dados disponíveis	Forma de obtenção
Q biogás	m ³ /dia	diária	942	medição
Q lamas	m ³ /dia	diária	941	medição
Q efluente bruto	m ³ /dia	diária	785	medição
T reator	°C	diária	906	medição
TRH	dias	diária	842	cálculo
AGV reator	mg/L	semanal	75	análise laboratorial
AGV/TAL	-	semanal	75	cálculo
CV	kg/(m ³ .dia)	semanal	108	cálculo
LMS in	kg/dia	semanal	100	cálculo
LMV in	kg/dia	semanal	108	cálculo
LMV out	kg/dia	semanal	112	cálculo
LMV removida	kg/dia	semanal	108	cálculo
MS in	mg/L	semanal	100	análise laboratorial
MS reator	mg/L	semanal	105	análise laboratorial
MV in	mg/L	semanal	108	análise laboratorial
MV reator	mg/L	semanal	112	análise laboratorial
pH in	-	semanal	82	análise laboratorial
pH reator	-	semanal	88	análise laboratorial
TAL reator	mg/L	semanal	75	análise laboratorial

Depois de concluída a recolha e acomodação dos dados em séries, optou-se por efetuar o cálculo do coeficiente de correlação de Pearson, o qual pode ser consultado na Tabela 5.19. No cálculo executado, foram excluídos todos os valores igual a zero, dado que estes se registam em dias onde ocorreram problemas operacionais e o digestor anaeróbio esteve parado. No entanto, foram considerados bastantes valores próximos de zero, os quais não se removeram por se achar importante efetuar o cálculo deste coeficiente com um teor de dados mais abrangente. O facto de não se ter retirado mais valores às séries de dados, permite ter uma melhor perceção da capacidade explicativa real de cada variável em relação aos valores de caudal de biogás produzido. A análise dos resultados do cálculo do coeficiente de Pearson é efetuada de acordo com os mesmos critérios dos da Tabela 5.4.

Tabela 5.19 – Resultados do cálculo do coeficiente de correlação de Pearson das séries de dados afetas ao segundo caso de estudo

	Q biogás	Q lamas	Q efluente bruto	T reator	TRH	AGV reator	AGV/TAL	CV	LMS in	LMV in	LMV out	LMV removida	MS in	MS reator	MV in	MV reator	pH in	pH reator	TAL reator
Q biogás																			
Q lamas	0,16																		
Q efluente	-0,01	0,14																	
T reator	0,45	0,06	-0,06																
TRH	-0,07	-0,2	-	-0,07															
AGV reator	-0,22	-0,09	-	-0,31	0,14														
AGV/TAL	0,23	0,02	-	-0,11	0,09	0,58													
CV	-0,1	-	-	-	-	-	-												
LMS in	0	-	-	-	-	-	-	-											
LMV in	-0,16	0,94	-	-0,1	-	-0,13	-	1	0,95										
LMV out	-0,24	0,73	-	-0,23	-0,45	0,14	-	-	-	0,65									
LMV removida	-0,01	0,91	-	-0,07	-0,18	-0,18	-0,12	-	-	0,98	0,52								
MS in	-0,15	-	-	-	-	-	-	-	-	-	-	-							
MS reator	-0,51	-	-	-	-	-	-	-	-	-	-	-	-						
MV in	0,28	-0,02	-	-0,17	0,13	-0,26	-	-	-	0,22	0,01	0,23	0,95	-					
MV reator	-0,48	-0,04	-	-0,51	-0,03	0,56	0,07	-	-	-0,04	0,5	-0,15	-	0,94	0,17				
pH in	0,09	0,11	-	0,17	-0,1	-0,21	0,18	-	0,03	-0,01	-	0,04	-	-	-0,34	-			
pH reator	0,64	0,07	-	0,55	-0,04	-0,53	-0,03	-	-	0,08	-	0,13	-	-	-0,14	-0,63	0,41		
TAL reator	-0,2	-0,26	-	0,21	0,05	-0,24	-0,83	-	-	-0,13	-0,13	-0,13	-	-	0,54	0,09	-0,31	-0,02	

Na observação dos resultados do cálculo do coeficiente de correlação de Pearson é possível verificar que existem nove correlações lineares fortes entre variáveis. Contudo, no que diz respeito às relações de correlação linear entre a variável “Q biogás” e as restantes, existem apenas quatro correlações moderadas e nenhuma forte. A variável “Q biogás” regista valores de correlação moderada com as variáveis “T reator”, “MS reator”, “MV reator” e “pH reator”. Apesar de o conjunto de variáveis disponível ser ligeiramente inferior ao existente na ETAR do Seixal, neste caso de estudo também se optou por reduzir o número de variáveis que viria a integrar os testes com as RNA. Esta decisão teve por base o facto de o tempo disponível para a execução da investigação ser inferior ao desejado, e o facto de o número de combinações de RNA possíveis de criar ser imenso, pelo que se tentou reduzir uma pequena percentagem das combinações possíveis. As cinco correlações lineares mais fortes observadas surgem entre as seguintes variáveis:

- “CV” e “LMV in”;
- “LMV in” e “LMV removida”;
- “LMV in” e “LMS in”;
- “MS in” e “MV in”;
- “MS reator” e “MV reator”.

Com base nas cinco correlações mais fortes verificadas e a correlação de cada uma das variáveis com a variável “Q biogás”, resolveu-se excluir as variáveis “CV”, “LMV removida”, “MS in” e “MS reator” dos testes a realizar com RNA. Nos testes a realizar deve-se dar prioridade às variáveis “T reator”, “MV reator” e “pH reator”, dado que, estas foram as variáveis que apresentaram uma correlação linear mais forte com os valores de caudal de biogás produzido.

Tendo em conta que a força das correlações verificadas, entre as diferentes séries de dados com a série de dados do caudal de biogás gerado não serem muito fortes, optou-se por usar alguns critérios de remoção de dados, de forma a ser possível testar diferentes configurações de dados e potenciar os resultados. Na seleção dos requisitos foi fundamental utilizar alguns dados estatísticos e executar vários gráficos afetos às variáveis consideradas. Os gráficos referentes às séries de dados de frequência de amostragem diária são apresentados em conjunto com os dados estatísticos, enquanto os restantes gráficos podem ser consultados no Anexo VII.

Sendo Sesimbra um destino de férias, de Verão, de muitas pessoas, quis-se perceber se este factor tinha alguma influência no caudal afluente à ETAR e também nos valores do caudal de lamas espessadas e de caudal de biogás gerado. A Tabela 5.20 apresenta as médias mensais de alguns dos parâmetros mais importantes.

Tabela 5.20 – Médias mensais de alguns parâmetros associados ao processo de digestão anaeróbia

	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Q lamas	15,69	16,68	12,35	17,17	17,56	22,70	23,67	33,46	23,70	11,95	15,21	9,66
T reator	17,58	26,23	24,39	25,97	27,13	28,40	30,07	29,57	26,14	29,60	26,15	18,66
Q ef. Bruto	2 588,48	2 680,44	3 344,35	2 822,86	2 444,56	2 474,78	2 811,44	3 026,44	2 501,07	2 271,76	2 515,00	2 255,40
Q biogás	274,67	266,11	135,36	264,33	228,53	212,84	220,93	251,09	210,59	165,29	165,29	97,48

Analisando a Tabela 5.20, a nível do caudal efluente bruto, as variações existentes são pequenas e não é possível identificar claramente uma época baixa e uma época alta. A temperatura no digestor, tal como na ETAR do Seixal, é um pouco mais elevada nos meses normalmente mais quentes em Portugal. A maior variação verificada ocorre nos valores dos caudais de lamas registados, os quais nos meses de verão são nitidamente mais elevados, contudo ao que tudo indica este facto pouco influencia a quantidade de biogás gerada, dado que neste parâmetro não existe um padrão sazonal. No global, as variações sazonais verificadas são pequenas e parecem influenciar pouco a produção de biogás.

Na Tabela 5.21 é possível consultar as principais estatísticas das séries de dados de frequência de amostragem diária, com a seguinte ressalva:

- O cálculo da média e do desvio padrão da série de dados da variável “Q biogás” e “Q lamas” foi efetuado na ausência dos valores igual a zero.

Tabela 5.21 – Estatísticas das séries de dados de frequência de amostragem diária

Variável	Unidade	Número de valores nulos	Valor mínimo	Valor máximo	Média	Desvio Padrão
Q biogás	m ³ /dia	47	0	2 811	135,62	207,57
Q lamas	m ³ /dia	99	0	78,40	15,19	12,90
Q efluente bruto	m ³ /dia	-	1 460	6 624	2664,53	638,11
T reator	°C	-	14,66	36,17	25,96	5,75
TRH	dias	-	7,70	10 000	265,65	936,63

Na análise da Tabela 5.21, destacam-se em primeiro lugar os valores estatísticos registados na série de dados da variável “TRH”, os quais são de todo anormais, no entanto deve-se ter em conta que estes resultam da divisão do volume do digestor pelos valores da variável “Q lamas”. Na série de dados do caudal de lamas espessadas que afluí ao digestor, foram considerados bastantes valores próximos de zero, pelo que estes dados produzem tempos de retenção enormes que desvirtuam a análise estatística da variável “TRH”.

A Temperatura do reator apesar de ser uma das séries com menos flutuações, é preciso notar que a média da temperatura registada está bastante afastada dos valores que se deviam praticar. Na Figura 5.9, pode-se observar a evolução da variável “T reator”.

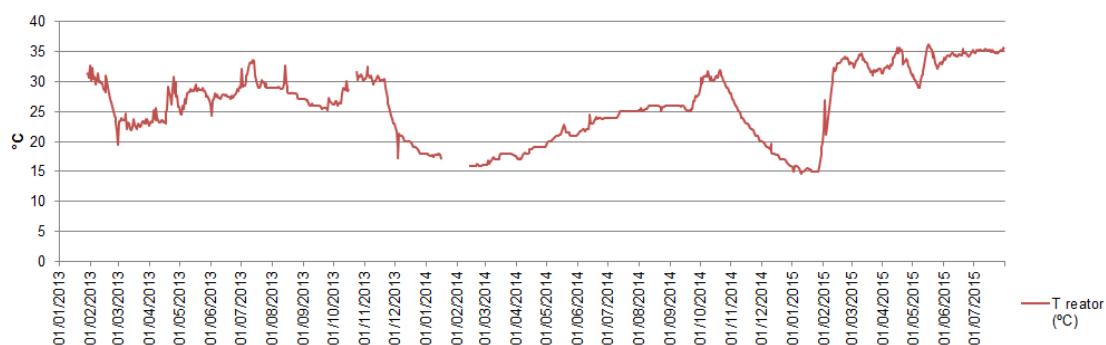


Figura 5.9 – Evolução da “T reator” – ETAR de Sesimbra

O processo de digestão anaeróbia deveria funcionar a uma temperatura mesófila, nomeadamente entre os 35°C e os 37°C (SIMARSUL, 2007). As temperaturas registadas, pelo facto de estarem afastadas dos valores ideais, podem ser um condicionante na eficiência do processo, ao afetar as taxas de atividade dos organismos bacterianos e consequentemente a produção de biogás.

Na Figura 5.10, é possível visualizar a evolução da variável “Q efluente bruto”.

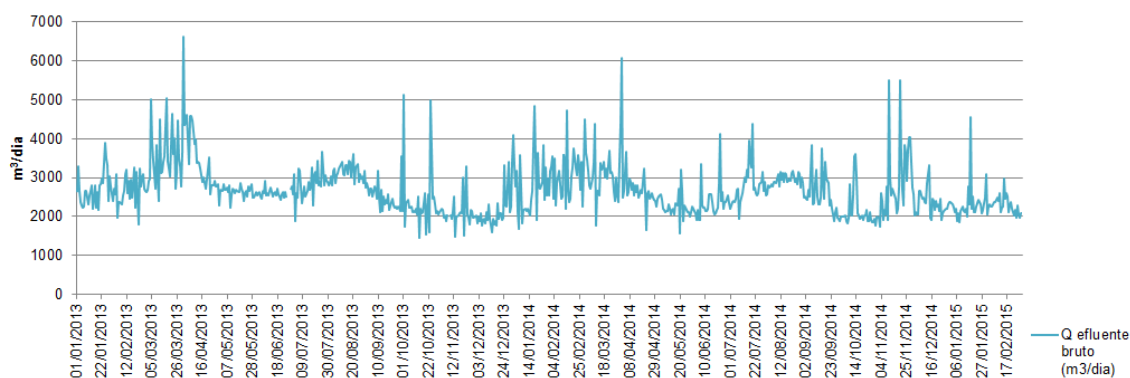


Figura 5.10 – Distribuição temporal da variável “Q efluente bruto” – ETAR de Sesimbra

A média dos valores de caudal afluente à ETAR de Sesimbra encontra-se bastante próxima do valor de dimensionamento para o ano zero (2 720 m³/dia), pelo que não se fez qualquer corte na série de dados. Em relação às séries de dados das variáveis “Q lamas” e “Q biogás”, estas apresentam alguma variância e valores que denotam problemas de operação em vários dias, pelo que em ambas se irá utilizar critérios de remoção de dados. Na Figura 5.11, mostra-se a evolução da variável “Q lamas”.

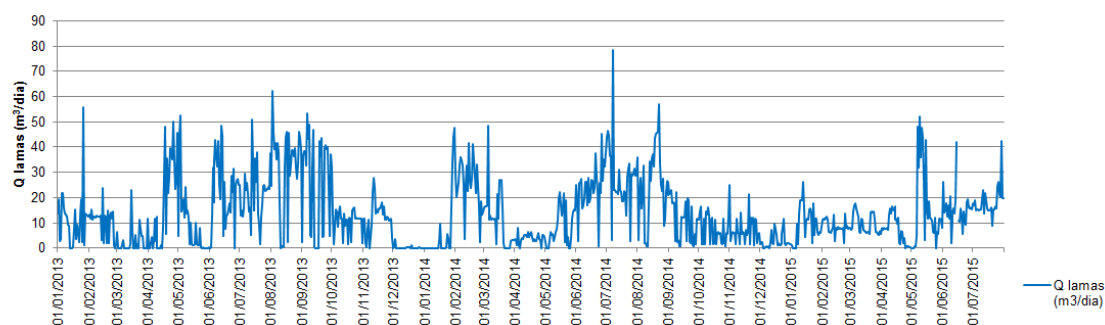


Figura 5.11 – Distribuição temporal das variáveis “Q lamas” – ETAR de Sesimbra

No caso da variável “Q lamas”, além dos valores nulos que nunca fazem parte das séries de dados que alimentam a produção de RNA, optou-se em muitos casos pela remoção de todos os valores abaixo dos 5 m³/dia. Na Tabela 5.22 é possível verificar a distribuição do número de dados da variável “Q lamas” por intervalo de valores.

Tabela 5.22 – Distribuição de dados da variável “Q lamas” por intervalo de valores

Número de valores nulos	Intervalo]0,5[Intervalo [5,15[Intervalo [15,30[Intervalo (>30)
99	200	307	220	115

Os valores do caudal de lamas espessadas afluentes ao digestor anaeróbio devem ser consideravelmente superiores a 5 m³/dia, contudo existem muitos dados abaixo deste valor, logo a remoção de um maior número de dados afetaria a significância dos resultados gerados. Decidiu-se nunca incluir os dados da variável “TRH” resultantes de valores de caudal de lamas espessadas inferiores a 5 m³/dia nas séries de dados utilizadas para desenvolvimento das RNA. A Tabela 5.23 apresenta as estatísticas da série de dados da variável “TRH” que foi usada nos testes com as RNA.

Tabela 5.23 – Estatísticas da série de dados da variável “TRH” utilizadas nos testes com as RNA

Valor mínimo	Valor máximo	Média	Desvio Padrão
7,70	117,20	44,30	26,60

Apesar da remoção dos maiores valores desta série de dados, ainda existem valores bastante elevados, o que resulta em uma média muito superior a 21 dias, valor de referência para o tempo de retenção de lamas no processo de digestão anaeróbia da ETAR de Sesimbra (SIMARSUL, 2007). A eliminação de mais dados afetos á variável “TRH” iria adulterar os resultados e diminuiria ainda mais o número de casos possíveis de destacar para as fases de treino e teste das redes neuronais, o que seria prejudicial. Em relação à série de dados do caudal de biogás gerado, são mostrados na Figura 5.12 a distribuição temporal do registo de

dados da variável “Q biogás” e na Tabela 5.24 a distribuição dos dados por intervalo de valores.

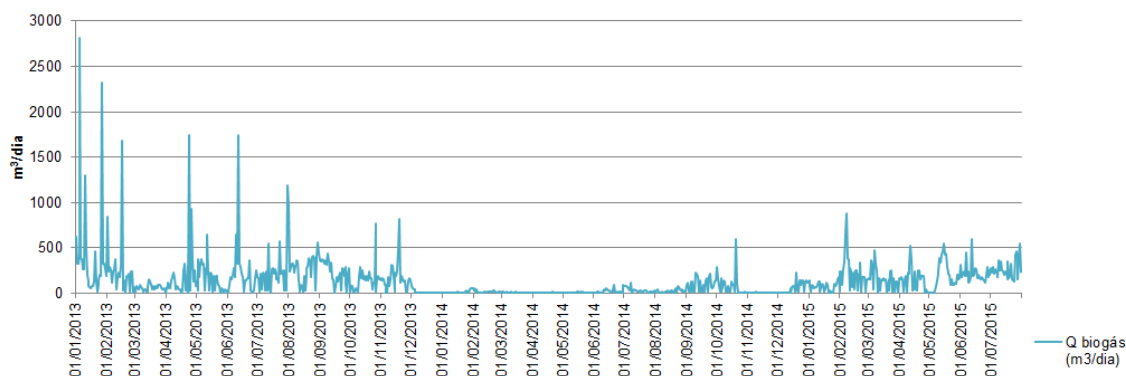


Figura 5.12 – Distribuição temporal da variável “Q biogás” – ETAR de Sesimbra

Tabela 5.24 – Distribuição de dados da variável “Q biogás” por intervalo de valores

Nº de valores nulos	Intervalo]0,10[Intervalo [10,20[Intervalo [20,50[Intervalo [50,100[Intervalo [100,200[Intervalo [200,400[Intervalo [400,600[Intervalo [600,1000[Intervalo (>1000)
47	204	95	99	102	180	173	25	10	7

Com base na análise estatística da série de dados completa do caudal de biogás, optou-se por não incluir os valores acima dos 1000 m³/dia no treino e teste das RNA. Os valores mais elevados estão associados a episódios de condensação no caudalímetro, levando a que estes sejam registados incorretamente, pelo que a sua inclusão poderia prejudicar a aprendizagem das RNA. Os valores acima de 600 m³/dia também são raros e ainda anormalmente elevados, porém considerou-se importante a sua manutenção, de forma a não se adulterar em demasia a base de dados. Em relação aos valores mais baixos verificados, estes estão associados a problemas operacionais mais recorrentes, pelo que se optou por eliminar um maior número de dados. Os problemas operacionais podem estar associados por exemplo à não alimentação de lamas ao digestor, devido a avarias na centrífuga ou a problemas no gasómetro que originam a saída de biogás pela válvula de sobrepressão do digestor, não passando pelo caudalímetro. Era difícil selecionar um intervalo de dados a remover, o que levou a que durante os testes com as RNA fossem considerados diferentes séries de dados da variável “Q biogás”. Na maioria das vezes foram considerados todos os valores de caudal de biogás acima de 0 m³/dia ou apenas acima de 20 m³/dia. Em outros casos consideraram-se os valores acima de 10 m³/dia ou 15 m³/dia.

No que diz respeito às séries de dados das variáveis de frequência de amostragem semanal, não se equacionou realizar qualquer corte na série de dados, visto o número de dados destas séries ser reduzido, pelo que qualquer corte diminuiria as hipóteses de produzir RNA com boas capacidades de predição de valores de biogás gerado. A Tabela 5.25 apresenta as estatísticas

das séries de dados das variáveis de frequência de amostragem semanal utilizadas no desenvolvimento das RNA, com a seguinte ressalva:

- A média e desvio padrão das série de dados das variáveis “LMV in” e “LMV out”, foram calculadas na ausência dos valores nulos.

Tabela 5.25 – Estatísticas das séries de dados de frequência de amostragem semanal

Variável	Unidade	Número de valores nulos	Valor mínimo	Valor máximo	Média	Desvio Padrão
AGV reator	mg/l	-	313	11 597	4 117	2 733
AGV/TAL	-	-	0,07	5,16	1,33	1,28
LMV in	Kg/dia	20	9	2 178	566	377
LMV out	Kg/dia	22	5	1 311	260	247
MV in	mg/l	-	8 367	77 000	33 979	14 591
MV reator	mg/l	-	3 133	35 850	15 284	7 260
pH in	-	-	4,90	6,72	5,51	0,32
pH reator	-	-	5,20	7,68	6,71	0,57
TAL reator	mg/l	-	774	8 442	4 444	1 747

Na Tabela 5.26 é efetuada a comparação das médias calculadas com os valores de referência.

Tabela 5.26 – Valores de referência para os parâmetros de frequência de amostragem semanal

Variável	Média calculada	Valores de referência
AGV reator	4 117	50 a 2000 (Malina e Pohland, 1992)
AGV/TAL	1,33	-
LMV in	566	-
LMV out	260	-
MV in	33 979	-
MV reator	15 284	-
pH in	5,51	5,0 a 7,5 (Qasim, 1999)
pH reator	6,71	6,4 a 7,4 (Turovskiy e Mathai, 2006)
TAL reator	4 444	2 000 a 5 000 (Metcalf & Eddy, 2003)

Segundo os valores de referência apurados, a maior parte dos parâmetros de carácter semanal estão dentro dos parâmetros normais de funcionamento do digestor anaeróbio, ao contrário do verificado com os parâmetros de frequência de amostragem diária. Todavia há que realçar que as séries de dados apresentam alguma variância de valores e que existem vários registos em determinados dias, que apresentam valores fora dos parâmetros normais de funcionamento de um digestor anaeróbio. As médias das variáveis “pH in” e “TAL reator” estão perto dos extremos dos valores de referência, mas ainda dentro dos valores de funcionamento normal. A média de valores do pH das lamas digeridas está dentro dos valores normais e muito próximo do intervalo ótimo, de 6,8 a 7,2 (Turovskiy e Mathai, 2006; Vesilind, 2003). Em relação aos

ácidos gordos voláteis (AGV), a sua média está fora dos valores normalmente encontrados, apesar de o funcionamento do digestor continuar a ser viável segundo os resultados apurados. No global, os valores afetos às variáveis utilizadas no desenvolvimento das RNA estão um pouco desviados dos parâmetros ótimos de funcionamento de um digestor anaeróbio. Registam-se inclusive problemas de exploração em bastantes dias, nomeadamente no ano de 2014, em que devido a problemas na etapa de desidratação, a digestão anaeróbia por vezes não pode funcionar normalmente. Os cortes efetuados nas séries de dados foram realizados com o intuito de dar mais consistência a estas mesmas séries, minimizando o impacto dos problemas operacionais verificados, tendo em consideração que no futuro estes problemas deverão ser menos comuns na instalação. Os critérios de remoção de dados usados, permitem gerar RNA com melhores capacidades de predição de valores de biogás gerado.

5.2.2 Etapa de Treino e Teste

A etapa de treino e teste realizada é um pouco complexa, pelo que antes de ser apresentada detalhadamente, é possível observar em esquema na Figura 5.13 o resumo desta fase da investigação.

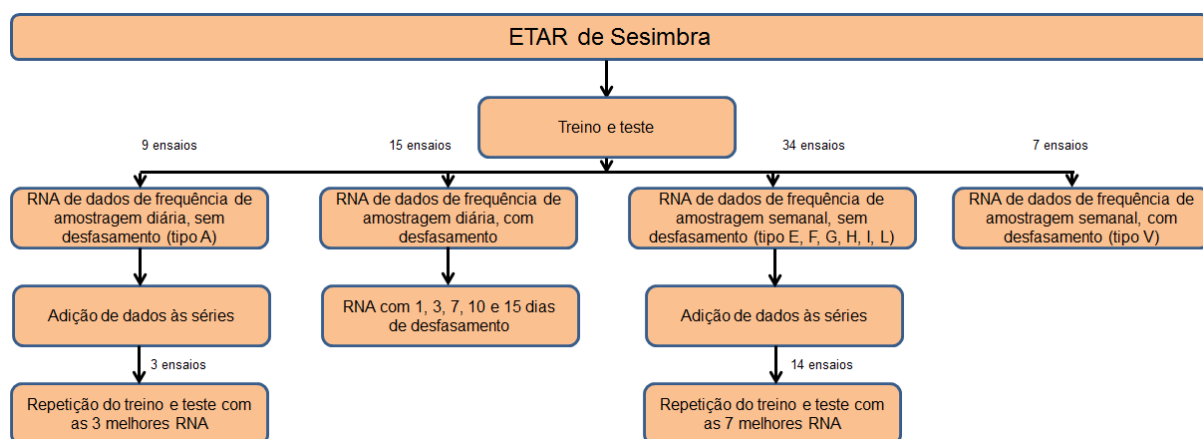


Figura 5.13 – Metodologia aplicada na etapa de treino e teste das redes neurais criadas com o intuito de prever valores de biogás produzido na ETAR de Sesimbra

Depois de concluído o tratamento de dados, iniciou-se a fase de treino e teste das RNA. Nesta etapa, as redes desenvolvidas foram divididas em quatro grupos:

- Redes neurais de séries de dados de frequência de amostragem diária, sem desfasamento de dados;
- Redes neurais de séries de dados de frequência de amostragem diária, com desfasamento de dados;
- Redes neurais de séries de dados de frequência de amostragem semanal, sem desfasamento de dados;
- Redes neurais de séries de dados de frequência de amostragem semanal, com desfasamento de dados.

Em cada grupo de RNA desenvolvido é possível distinguir dois períodos temporais, no primeiro as redes neuronais foram desenvolvidas com base em dados de 1 de Janeiro de 2013 a 23 de Abril de 2015. No final deste período, de forma a se selecionar as melhores redes neuronais produzidas até ao momento, utilizaram-se os seguintes critérios:

- $\leq 50\%$ de más previsões (30% de tolerância) nos casos destacados para treino;
- $\leq 50\%$ de más previsões (30% de tolerância) nos casos destacados para teste;
- Erro médio quadrático ≤ 50 nos casos destacados para treino;
- Erro médio quadrático ≤ 75 nos casos destacados para teste.

Só as RNA que satisfaziam pelo menos três destas condições eram identificadas como redes neuronais com bom potencial de previsão de biogás produzido e consideradas para novos testes. No segundo período, as RNA foram desenvolvidas com base em dados de 1 de Janeiro de 2015 a 31 de Julho de 2015. A adição de dados que ocorreu foi executada com o objetivo de aumentar o número de casos de treino e teste, de modo a tentar melhorar os resultados das melhores redes identificadas durante o primeiro período.

5.2.2.1 Redes neuronais de séries de dados de frequência de amostragem diária, sem desfasamento de dados

Durante a investigação foram desenvolvidas 12 RNA afetas a este grupo, sendo que três destas 12 foram produzidas a partir das séries de dados de maior extensão temporal disponíveis. Das nove redes iniciais produzidas, nenhuma cumpriu três dos quatro critérios definidos, na verdade apenas uma das RNA cumpriu dois dos requisitos, pelo que à partida nenhuma rede foi identificada como tendo um bom potencial de previsão de valores de biogás. Contudo como se considerava importante para o estudo utilizar pelo menos uma RNA de frequência de amostragem diária na previsão de valores de biogás, selecionaram-se três RNA para novos testes. Tendo em conta os resultados verificados, as três redes neuronais foram selecionadas por possuírem um erro médio quadrático no treino e teste inferior a 130, o que as destacou claramente de todas as outras redes criadas neste grupo. Na Tabela 5.27 é possível constatar os resultados das três redes selecionadas, designadas por “1A1”, “2A1” e “3A1”.

Tabela 5.27 – Resultados das redes neuronais “1A1”, “2A1” e “3A1”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
1A1	TRH (dias)	38,00%	445	01/01/2013	23/04/2015	Melhor rede	MLFN (10 nós)
	Q lamas (m ³ /dia)	37,29%				% más previsões (treino - 30% toler.)	49,16%
	T reator (°C)	24,71%				% más previsões (teste - 30% toler.)	43,82%
						Erro médio quadrático (treino)	100,51
						Erro médio quadrático (teste)	124,84
2A1	T reator (°C)	56,02%	445	01/01/2013	23/04/2015	Melhor rede	MLFN (6 nós)
	Q lamas (m ³ /dia)	43,98%				% más previsões (treino - 30% toler.)	50,56%
						% más previsões (teste - 30% toler.)	57,30%
						Erro médio quadrático (treino)	114,33
						Erro médio quadrático (teste)	86,48
3A1	T reator (°C)	35,06%	391	01/01/2013	01/03/2015	Melhor rede	MLFN (3 nós)
	Q lamas (m ³ /dia)	22,92%				% más previsões (treino - 30% toler.)	53,04%
	Q efluente bruto (m ³ /dia)	21,22%				% más previsões (teste - 30% toler.)	58,97%
	TRH (dias)	20,80%				Erro médio quadrático (treino)	109,81
						Erro médio quadrático (teste)	120,67

Os resultados das três redes neuronais exibidas na Tabela 5.27 foram gerados, tendo por base séries de dados em que foram removidos todos os valores da variável “Q lamas” abaixo dos 5 m³/dia e todos os valores da variável “Q biogás” acima de 1000 m³/dia e abaixo de 20 m³/dia. A rede “3A1” possui dados apenas até 1 de Março de 2013, visto que o número de dados recolhidos da variável “Q efluente bruto” foi de menor número. Consumada a adição de dados às séries, foram realizados novos testes com as configurações das redes “1A1”, “2A1” e “3A1”, os resultados podem ser observados na Tabela 5.28.

Tabela 5.28 – Resultados dos novos testes realizados com as configurações das redes “1A1”, “2A1” e “3A1”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
1A1 2V	TRH (dias)	39,25%	527	01/01/2013	31/07/2015	Melhor rede	MLFN (8 nós)
	Q lamas (m ³ /dia)	37,24%				% más previsões (treino - 30% toler.)	47,16%
	T reator (°C)	23,50%				% más previsões (teste - 30% toler.)	52,38%
						Erro médio quadrático (treino)	105,86
						Erro médio quadrático (teste)	105,43
2A1 2V	T reator (°C)	52,32%	527	01/01/2013	31/07/2015	Melhor rede	MLFN (4 nós)
	Q lamas (m ³ /dia)	47,68%				% más previsões (treino - 30% toler.)	50,95%
						% más previsões (teste - 30% toler.)	57,14%
						Erro médio quadrático (treino)	111,94
						Erro médio quadrático (teste)	107,54
3A1 2V	Q lamas (m ³ /dia)	31,71%	391	01/01/2013	01/03/2015	Melhor rede	MLFN (5 nós)
	T reator (°C)	30,38%				% más previsões (treino - 30% toler.)	49,84%
	TRH (dias)	24,64%				% más previsões (teste - 30% toler.)	60,26%
	Q efluente bruto (m ³ /dia)	13,27%				Erro médio quadrático (treino)	98,23
						Erro médio quadrático (teste)	120,40

A adição de dados não teve grande impacto nos resultados registados. De maneira geral as alterações verificadas nos resultados não foram significativas e considera-se que os novos resultados são do mesmo nível que os anteriores. A razão mais provável para esta situação é o facto de o número de dados adicionado ser pouco relevante face ao tamanho da série de dados existente, a qual já era bastante extensa. Os dados adicionados pouco ou nada

mudaram as características das séries de dados, pelo que os resultados gerados foram muito semelhantes aos anteriores.

5.2.2.2 Redes neuronais de séries de dados de frequência de amostragem diária, com desfasamento de dados

Neste grupo de RNA, foram testadas as três configurações de séries de dados que melhores resultados produziram no grupo anterior de redes neuronais desenvolvidas, ou seja as configurações das redes “1A1”, “2A1” e “3A1”. Apesar de as configurações serem as mesmas, neste caso, optou-se por desfasar a variável “Q biogás” das restantes variáveis em 1 dia, 3 dias, 7 dias, 10 dias e por último em 15 dias. No primeiro período de desenvolvimento das RNA deste grupo foram geradas 15 RNA, contudo quando se realizou a análise final dos resultados gerados, antes de ocorrer a adição de dados, descobriu-se alguns problemas. As séries de dados que alimentaram a produção destas redes continham gralhas, pelo que os resultados gerados por estas RNA foram de pronto anulados. Em seguida, decidiu-se corrigir os erros encontrados, adicionar os novos dados disponibilizados e reproduzir de novo todas as 15 RNA criadas anteriormente, o que atrasou um pouco a restante investigação por executar. Concluído o desenvolvimento das RNA, realizou-se a análise dos resultados, onde se constatou que nenhuma das redes criadas cumpria os requisitos, de modo a ser identificada como uma rede neuronal com bom potencial de predição de valores de biogás. A Tabela 5.29 exhibe a média dos resultados gerados por número de dias de desfasamento, enquanto a Figura 5.14 expõe a representação gráfica dessas mesmas médias. Na Tabela 5.29 e Figura 5.14 é também acrescentada a média dos resultados das redes “1A1 2V”, “2A1 2V” e “3A1 2V”, de forma a ser possível comparar os resultados entre as RNA sem desfasamento de dados com as RNA com desfasamento de dados.

Tabela 5.29 – Média dos resultados gerados por número de dias de desfasamento

Tipo de Redes	Resultados	
sem desfasamento	% más previsões (treino - 30% toler.)	49,32%
	% más previsões (teste - 30% toler.)	56,59%
	Erro médio quadrático (treino)	105,34
	Erro médio quadrático (teste)	111,12
1 dia de desfasamento	% más previsões (treino - 30% toler.)	56,92%
	% más previsões (teste - 30% toler.)	57,53%
	Erro médio quadrático (treino)	148,19
	Erro médio quadrático (teste)	166,18
3 dias de desfasamento	% más previsões (treino - 30% toler.)	59,02%
	% más previsões (teste - 30% toler.)	60,55%
	Erro médio quadrático (treino)	160,21
	Erro médio quadrático (teste)	167,14
7 dias de desfasamento	% más previsões (treino - 30% toler.)	53,40%
	% más previsões (teste - 30% toler.)	60,49%
	Erro médio quadrático (treino)	153,49
	Erro médio quadrático (teste)	219,37
10 dias de desfasamento	% más previsões (treino - 30% toler.)	61,94%
	% más previsões (teste - 30% toler.)	65,38%
	Erro médio quadrático (treino)	156,04
	Erro médio quadrático (teste)	144,64
15 dias de desfasamento	% más previsões (treino - 30% toler.)	58,84%
	% más previsões (teste - 30% toler.)	53,76%
	Erro médio quadrático (treino)	148,01
	Erro médio quadrático (teste)	162,02

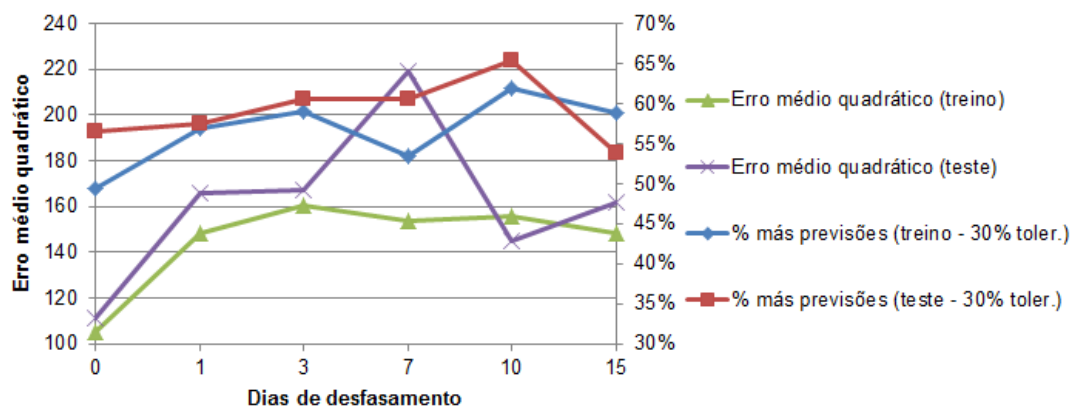


Figura 5.14 – Média dos resultados gerados por número de dias de desfasamento

A Tabela 5.29 e a Figura 5.14 demonstram claramente que as RNA sem desfasamento de dados produziram melhores resultados. A nível das RNA com desfasamento de dados, é difícil identificar qual o intervalo de desfasamento que produziu melhores resultados, dado os

resultados variarem consoante o parâmetro considerado. Se considerarmos só o erro médio quadrático verificado durante os testes, o qual é um parâmetro bastante importante para identificar a capacidade de previsão de uma rede, a nível de desfasamento de dados, 10 dias seria o intervalo de tempo que apresentou um melhor resultado. Avaliando a performance a nível geral, o intervalo de desfasamento entre dados de 15 dias, foi o único que apresentou melhores resultados em duas categorias, pelo que este seria o melhor intervalo. Porém, neste estudo, é pouco relevante identificar o melhor intervalo de desfasamento a considerar, visto que os resultados entre intervalos foram idênticos e em ambos os casos inferiores aos resultados gerados sem desfasamento de dados.

5.2.2.3 Redes neuronais de séries de dados de frequência de amostragem semanal, sem desfasamento de dados

O maior número de RNA treinadas e testadas tiveram por base séries de dados de frequência de amostragem semanal, sem desfasamento de dados. No primeiro período temporal de testes, foram desenvolvidos 34 redes neuronais, das quais 17 cumpriram os requisitos de modo a serem consideradas RNA de bom potencial de previsão, ou seja, exatamente 50% das redes criadas. A Tabela 5.30 mostra o número de utilizações de cada variável nas RNA produzidas e o número de vezes que cada variável está presente nas redes que cumpriram os critérios estabelecidos.

Tabela 5.30 – Número de utilizações de cada variável na produção das redes neuronais

Variável	Número de presenças no total de redes neuronais geradas	Número de presenças nas melhores redes neuronais geradas	Taxa de sucesso
Q lamas	12	7	58,33%
Q efluente bruto	5	4	80,00%
T reator	34	17	50,00%
TRH	18	10	55,56%
AGV reator	18	10	55,56%
AGV/TAL	8	4	50,00%
LMV in	11	8	72,73%
LMV out	5	4	80,00%
MV in	9	5	55,56%
MV reator	19	10	52,63%
pH in	4	3	75,00%
pH reator	33	17	51,52%
TAL reator	13	6	46,15%

Na análise à Tabela 5.30 é possível constatar que as variáveis mais utilizadas foram “T reator”, “MV reator” e “pH reator”. A nível da taxa de sucesso, não é possível concluir acerca das

variáveis que mostraram uma melhor performance, dado que os coeficientes são muito idênticos. As variáveis “Q efluente bruto” e “LMV out” apresentam as taxas mais elevadas, no entanto também foram das variáveis menos usadas, pelo que o resultado apresentado pode ser enganoso. O facto de a variável “T reator” ter estado em todas as configurações e a variável “pH reator” em quase todas, também complica a análise de resultados. Estas variáveis pela sua presença nas configurações podem permitir que outras variáveis menos explicativas dos valores de caudal de biogás produzido, também marquem presença nas configurações que apresentaram melhores resultados. Se o tempo disponível para o período de testes tivesse sido maior, dever-se-ia ter estudado mais configurações, inclusive configurações em que as variáveis que apresentaram melhores valores de coeficiente de correlação com a variável “Q biogás” não estivessem presentes.

Nas 17 RNA que satisfizeram três dos quatro critérios definidos, encontram-se as configurações ou configurações aproximadas das cinco melhores redes neuronais produzidas no caso de estudo da ETAR do Seixal. Em relação aos valores de caudal de biogás das séries de dados que serviram de base à produção destas 17 RNA, em oito foram considerados valores acima dos $20\text{m}^3/\text{dia}$, enquanto em outras oito foram considerados valores acima dos $0\text{m}^3/\text{dia}$. Em dois casos foram considerados valores acima de $10\text{m}^3/\text{dia}$ e $15\text{m}^3/\text{dia}$ (estas opções foram menos usadas que as anteriores). Contudo testar novamente 17 RNA com uma base de dados alargada, era uma abordagem que levaria bastante tempo a concluir, pelo que se decidiu encurtar esse número. A opção tomada levou a que no caso das redes neuronais de frequência de amostragem semanal sem desfasamento, se considerassem como melhores redes, apenas aquelas que cumpriram os quatro requisitos definidos ou contivessem um erro médio quadrático na fase de testes inferior a 50. Dentro dos novos critérios selecionados enquadram-se sete RNA, sendo que os seus resultados podem ser observados na Tabela 5.31.

Tabela 5.31 - Resultados das melhores redes neurais geradas a partir de séries de dados de frequência de amostragem semanal, sem desfasamento de dados

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
2E	T reator (°C)	55,92%	52	05/02/2013	23/04/2015	Melhor rede	GRNN
	pH reator	44,08%				% más previsões (treino - 30% toler.)	40,48%
	MV reator (mg/L)	0,00%				% más previsões (teste - 30% toler.)	40,00%
						Erro médio quadrático (treino)	46,75
						Erro médio quadrático (teste)	47,38
4E	pH reator	19,32%	34	05/02/2013	26/02/2015	Melhor rede	MLFN (12 nós)
	TRH (dias)	18,93%				% más previsões (treino - 30% toler.)	0,00%
	Q efluente bruto (m³/dia)	17,39%				% más previsões (teste - 30% toler.)	50,00%
	MV reator (mg/L)	16,05%				Erro médio quadrático (treino)	0,01
	T reator (°C)	15,91%				Erro médio quadrático (teste)	69,52
	Q lamas (m³/dia)	12,40%					
3I	MV in (mg/L)	21,52%	32	05/02/2013	23/04/2015	Melhor rede	MLFN (6 nós)
	AGV/TAL	20,13%				% más previsões (treino - 30% toler.)	0,00%
	pH reator	18,09%				% más previsões (teste - 30% toler.)	50,00%
	MV reator (mg/L)	15,95%				Erro médio quadrático (treino)	0,01
	LMV out (kg/dia)	13,41%				Erro médio quadrático (teste)	71,71
	T reator (°C)	10,89%					
2L	MV in (mg/L)	43,45%	21	05/02/2013	26/02/2015	Melhor rede	GRNN
	AGV reator (mg/L)	25,39%				% más previsões (treino - 30% toler.)	29,41%
	pH reator	18,73%				% más previsões (teste - 30% toler.)	25,00%
	T reator (°C)	10,91%				Erro médio quadrático (treino)	29,15
	LMV in (kg/dia)	0,53%				Erro médio quadrático (teste)	34,95
	LMV out (kg/dia)	0,32%					
	AGV/TAL	0,17%					
	Q lamas (m³/dia)	0,16%					
	MV reator (mg/L)	0,13%					
	pH in	0,13%					
	TAL reator (mg/L)	0,05%					
	TRH (dias)	0,03%					
	Q efluente bruto (m³/dia)	0,00%					
8F sx	AGV reator (mg/L)	49,24%	36	05/02/2013	23/04/2015	Melhor rede	GRNN
	MV reator (mg/L)	20,83%				% más previsões (treino - 30% toler.)	24,14%
	pH reator	18,67%				% más previsões (teste - 30% toler.)	42,86%
	T reator (°C)	10,21%				Erro médio quadrático (treino)	25,17
	TRH (dias)	0,58%				Erro médio quadrático (teste)	61,43
	TAL reator (mg/L)	0,47%					
2 8H sx	T reator (°C)	28,34%	54	05/02/2013	23/04/2015	Melhor rede	MLFN (7 nós)
	TAL reator (mg/L)	27,67%				% más previsões (treino - 30% toler.)	11,63%
	AGV reator (mg/L)	23,56%				% más previsões (teste - 30% toler.)	54,55%
	pH reator	20,42%				Erro médio quadrático (treino)	3,40
						Erro médio quadrático (teste)	45,28
1 9G8 sx	pH reator	71,34%	39	05/02/2013	23/04/2015	Melhor rede	GRNN
	T reator (°C)	16,36%				% más previsões (treino - 30% toler.)	48,28%
	TRH (dias)	12,26%				% más previsões (teste - 30% toler.)	40,00%
	LMV in (kg/dia)	0,04%				Erro médio quadrático (treino)	55,94
						Erro médio quadrático (teste)	64,36

As configurações das últimas três redes neurais exibidas na Tabela 5.31 são inspiradas nas RNA que exibiram melhores resultados na ETAR do Seixal, no caso da rede “8F sx”, a configuração é exatamente igual. Em relação a estas sete redes neurais, as redes neurais “8F sx” e a “2 8H sx” foram criadas a partir de séries de dados de caudal de biogás superiores a 15 m³/dia e 0 m³/dia, respetivamente. As restantes só contemplam valores acima dos 20 m³/dia. Os resultados no geral são semelhantes, mas há que destacar o erro médio quadrático na etapa de teste registado na rede “2L”, o qual é bastante baixo, no entanto esta rede foi criada com pouquíssimos casos levando a que nesta fase não se tenha a certeza se os

resultados gerados na previsão serão da mesma qualidade. A Tabela 5.32 é idêntica à 5.30 e apresenta o número de presenças de cada variável nas melhores sete RNA desenvolvidas.

Tabela 5.32 – Número de presenças de cada variável nas configurações das melhores redes neurais criadas

Variável	Número de presenças nas melhores dezassete redes neurais geradas	Número de presenças nas melhores sete redes neurais geradas	Taxa de sucesso
Q lamas	7	2	28,57%
Q efluente bruto	4	2	50,00%
T reator	17	7	41,18%
TRH	10	4	40,00%
AGV reator	10	3	30,00%
AGV/TAL	4	2	50,00%
LMV in	8	2	25,00%
LMV out	4	2	50,00%
MV in	5	2	40,00%
MV reator	10	5	50,00%
pH in	3	1	33,33%
pH reator	17	7	41,18%
TAL reator	6	3	50,00%

As variáveis com maior número de presenças nas configurações das sete melhores RNA criadas são naturalmente as variáveis “T reator” e “pH reator” dado que também foram as mais utilizadas. Analisando a Tabela 5.32 no geral e o impacto que cada variável demonstra ter nos resultados (Tabela 5.31), é possível concluir que todas as variáveis usadas no treino e teste das redes neurais têm importância para o estudo e contribuíram para a aprendizagem das RNA desenvolvidas.

Executada a análise às redes neurais criadas, adicionaram-se novos dados às séries de dados de frequência de amostragem semanal, com exceção à rede “2L”, e de seguida, voltaram-se a treinar e a testar RNA com base nas configurações que geraram os melhores resultados. Foram treinadas e testadas dois grupos de redes neurais por cada uma das sete configurações selecionadas, de modo a conferir se era possível produzir melhores resultados com as novas séries de dados. Os resultados anteriores e os novos resultados gerados foram analisados para cada configuração, de forma a se definir qual a melhor rede gerada de cada configuração, segundo os seguintes critérios:

- O primeiro critério de definição é o número de requisitos anteriormente apresentados, que cada rede neuronal cumpre;
- O segundo critério de definição é o valor do erro médio quadrático da fase de teste de cada rede neuronal.

Esperava-se que com o aumento do número de dados, os resultados melhorassem, contudo essa presunção apenas se verificou em um caso, nomeadamente na rede “1 9G8 sx”. O resultado afeto à nova melhor rede “1 9G8 sx” é apresentado na Tabela 5.33.

Tabela 5.33 – Resultado da nova melhor rede “1 9G8 sx”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
1 9G8 sx	T reator (°C)	90,48%	49	05/02/2013	30/07/2015	Melhor rede	GRNN
	pH reator	9,43%				% más previsões (treino - 30% toler.)	35,14%
	TRH (dias)	0,06%				% más previsões (teste - 30% toler.)	50,00%
	LMV in (kg/dia)	0,02%				Erro médio quadrático (treino)	54,43
						Erro médio quadrático (teste)	57,14

Conforme é passível de observação na Tabela 5.33, os últimos testes executados permitiram produzir uma nova rede “1 9G8 sx” que cumpriu os quatro requisitos apresentados no início deste subcapítulo e detém um erro médio quadrático na fase de teste inferior ao verificado anteriormente.

5.2.2.4 Redes neuronais de séries de dados de frequência de amostragem semanal, com desfasamento de dados

As configurações das redes neuronais afetas a este grupo foram as mais difíceis de definir. Quando se fala de variáveis de frequência de amostragem semanal, significa que todas as semanas é registado um valor afeto a cada variável, no entanto o intervalo entre medições não é necessariamente de sete dias. A Tabela 5.34 expõe a frequência com que cada intervalo de um determinado número de dias ocorreu entre amostragens.

Tabela 5.34 – Frequência de intervalos entre amostragens

Intervalo de dias	Frequência
4	3
5	18
6	14
7	34
8	59
9	3
12	2

O método mais correto para configurar as séries de dados deste grupo de RNA, seria a introdução dos valores desfasados de caudal de biogás diretamente no conjunto de dados. Na Tabela 5.34 é possível confirmar que a maior parte de amostragens ocorreu de oito em oito dias, todavia registam-se muitos outros intervalos de amostragem, o que complicava a decisão a tomar sobre o desfasamento de dias a aplicar. A diferença temporal dos períodos entre amostragens e o tempo necessário a despender na criação das séries de dados, impugnou a abordagem mais correta a realizar, tal como se verificou na investigação executada na ETAR

do Seixal no ano de 2014. A opção tomada neste estudo foi igual à selecionada na interior investigação, ou seja, optou-se por recuar uma célula os valores referentes à variável “Q biogás”.

As redes neuronais afetas a este grupo foram as últimas a ser desenvolvidas, pelo que no momento do início da sua criação, a última remessa de dados tinha acabado de ser recolhida, levando a que todos os dados se encontrassem disponíveis. No entanto, as redes neuronais geradas a partir das séries completas de dados de frequência de amostragem semanal sem desfasamento, não tinham ainda sido desenvolvidas, pelo que não se tinha conhecimento de que a adição de dados não tinha melhorado os resultados. À semelhança da opção tomada em relação ao grupo de redes neuronais de frequência de amostragem diária com desfasamento, no grupo atual também se decidiu treinar e testar RNA, com base nas configurações que melhores resultados proporcionaram através da utilização de séries de dados sem desfasamento. Neste grupo optou-se desde as primeiras RNA geradas, por se utilizar as séries de dados completas, dada a existência desta possibilidade. A Tabela 5.35 permite a observação dos resultados das melhores RNA, concebidas com base em séries de dados de frequência de amostragem semanal com desfasamento entre dados.

Tabela 5.35 – Melhores redes neuronais de séries de dados de frequência de amostragem semanal, com desfasamento entre dados

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
3V	pH reator	31,73%	42	14/02/2013	30/07/2015	Melhor rede	GRNN
	MV reator (mg/L)	21,08%				% más previsões (treino - 30% toler.)	0,00%
	AGV reator (mg/L)	19,76%				% más previsões (teste - 30% toler.)	50,00%
	TRH (dias)	10,77%				Erro médio quadrático (treino)	3,44
	TAL reator (mg/L)	10,35%				Erro médio quadrático (teste)	72,97
	T reator (°C)	6,31%					
4V	Q efluente bruto (m³/dia)	29,62%	35	14/02/2013	03/03/2015	Melhor rede	MLFN (3 nós)
	MV reator (mg/L)	28,30%				% más previsões (treino - 30% toler.)	19,23%
	pH reator	21,75%				% más previsões (teste - 30% toler.)	44,44%
	T reator (°C)	9,77%				Erro médio quadrático (treino)	21,68
	Q lamas (m³/dia)	5,48%				Erro médio quadrático (teste)	72,24
	TRH (dias)	5,08%					
6V	pH reator	14,25%	18	14/02/2013	03/03/2015	Melhor rede	MLFN (13 nós)
	LMV in (kg/dia)	9,24%				% más previsões (treino - 30% toler.)	0,00%
	Q lamas (m³/dia)	9,19%				% más previsões (teste - 30% toler.)	50,00%
	T reator (°C)	8,64%				Erro médio quadrático (treino)	0,00
	pH in	8,46%				Erro médio quadrático (teste)	57,18
	MV reator (mg/L)	7,80%					
	Q efluente bruto (m³/dia)	7,72%					
	TRH (dias)	7,53%					
	MV in (mg/L)	6,14%					
	AGV reator (mg/L)	5,75%					
	TAL reator (mg/L)	5,73%					
	LMV out (kg/dia)	5,16%					
	AGV/TAL	4,39%					

Das sete redes neuronais geradas, apenas três cumpriram pelo menos três dos requisitos definidos para serem consideradas RNA com potencial de previsão. As configurações de variáveis das redes neuronais “3V”, “4V” e “6V” são respetivamente iguais às das redes “8F sx”,

“4E” e “2L”. No geral os resultados gerados pelas séries de dados com desfasamento foram piores, inclusive os resultados das redes da Tabela 5.35. As melhorias verificadas dizem respeito apenas às fases de treino das redes “3V” e “6V”.

5.2.3 Etapa de Previsão

De forma a simplificar a compreensão das ações tomadas nesta etapa, é apresentado em esquema na Figura 5.15 a metodologia aplicada.

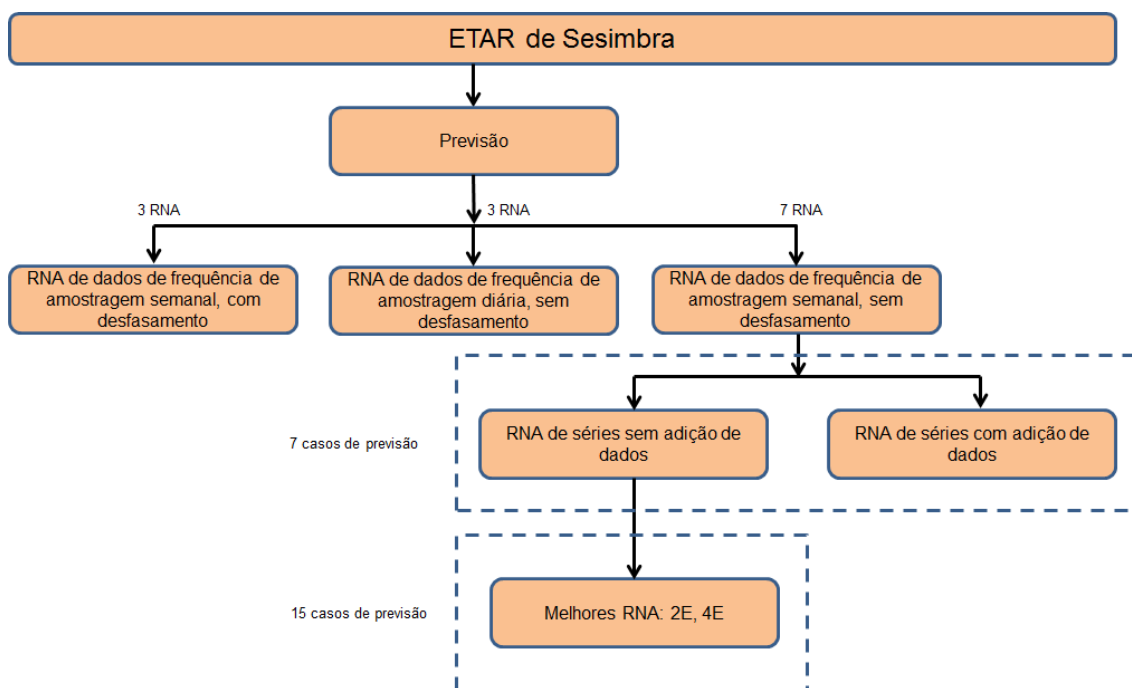


Figura 5.15 – Metodologia aplicada na etapa de previsão na ETAR de Sesimbra

Concluído o desenvolvimento de todas as RNA e respetivas análises de resultados das fases de treino e teste, foi possível iniciar a etapa de previsão dos valores de caudal de biogás. Na etapa de previsão foram utilizadas as seguintes RNA:

- Redes neuronais de séries de dados de frequência de amostragem semanal, com desfasamento de dados, nomeadamente as redes “3V”, “4V” e “6V”;
- Redes neuronais de séries de dados de frequência de amostragem diária, sem desfasamento de dados, nomeadamente as redes “1A1 2V”, “2A1 2V” e “3A1 2V”;
- Redes neuronais de séries de dados de frequência de amostragem semanal, sem desfasamento de dados, nomeadamente as redes “2E”, “4E”, “3I”, “2L”, “8F sx”, “2 8H sx” e “1 9G8 sx”.

O melhor resultado da fase de previsão das redes neuronais de séries de dados de frequência de amostragem semanal, com desfasamento de dados, pode ser consultado na Tabela 5.36. Os casos de previsão datam de 12 de Agosto de 2015 a 24 de Setembro de 2015.

Tabela 5.36 – Melhor resultado na fase de previsão das redes neuronais com desfasamento de dados

Rede	Casos treino + teste	Casos previsão	RMS	R ²
6V	18	7	125,04	0,37

A análise à Tabela 5.36 à primeira vista pode indicar que o resultado da rede neuronal “6V” é razoável, do ponto de vista global da investigação, face aos restantes resultados da etapa de previsão disponibilizados. Porém, o facto de terem sido utilizados apenas sete casos para previsão pode estar a inflacionar o resultado, que deverá piorar quando a fase de previsão for efetuada com um maior número de casos destacados. Em relação às restantes duas RNA afetas a esta categoria de redes destacadas para a etapa de previsão, os resultados foram bastantes inferiores ao da rede “6V”. Estes resultados vêm confirmar o indício dado pelos resultados da fase de treino e teste, os quais indicavam a rede “6V” como a que tinha maior potencial para a predição de valores de biogás, apesar de ter sido a configuração treinada com um menor número de dados.

Os resultados mais interessantes da fase de previsão das redes neuronais de séries de dados de frequência de amostragem diária, sem desfasamento de dados, podem ser consultados na Tabela 5.37. Os casos de previsão datam de 1 de Agosto de 2015 a 29 de Setembro de 2015.

Tabela 5.37 – Melhores resultados na fase de previsão das redes neuronais com dados de frequência de amostragem diária

Rede	Casos treino + teste	Casos previsão	RMS	R ²
1A1 2V	527	59	125,99	0,12
2A1 2V	527	59	118,17	0,14

No desenvolvimento das redes neuronais de dados de frequência de amostragem diária, sem desfasamento, foram produzidas duas RNA de capacidade de previsão bastante idêntica. A fase de treino e teste indicava que a rede “1A1 2V” podia ser ligeiramente melhor, todavia a rede “2A1 2V” foi a que apresentou melhores resultados na fase de previsão. A rede “3A1 2V” tal como na fase de treino e teste voltou a apresentar resultados inferiores. Os resultados expostos na Tabela 5.37 podem indicar que perante resultados nas fases de treino e teste bastante semelhantes, nem sempre a rede neuronal que apresenta melhores resultados nessa etapa é também a que apresenta melhores resultados na fase de previsão.

No que diz respeito à etapa de previsão das RNA de dados de frequência de amostragem semanal, sem desfasamento, a abordagem foi um pouco diferente. Optou-se por se conferir se os resultados ligeiramente piores, verificados após a adição de casos de treino e teste às séries das sete melhores RNA deste grupo, significam de facto que na fase de previsão também o seriam. Esta comparação teve por base a execução da etapa de previsão para

apenas sete casos entre as datas de 12 de Agosto de 2015 a 24 de Setembro de 2015, de modo a que esta fase ocorresse nas mesmas condições para todas as RNA. A Tabela 5.38 exhibe os resultados alcançados.

Tabela 5.38 – Resultados da fase de previsão das redes neuronais de dados de frequência de amostragem semanal, sem desfasamento

Rede	Séries com ou sem adição de dados	Casos treino + teste	Casos previsão	RMS	R ²
2E	sem adição	52	7	80,89	0,62
4E	sem adição	34	7	41,79	0,81
3I	com adição	41	7	29,77	0,80
2L	-	21	7	91,44	0,15
8F sx	sem adição	36	7	127,76	0,41
2 8H sx	com adição	69	7	58,73	0,24
1 9G8 sx	com adição	49	7	35,94	0,80

Os resultados da fase de previsão da Tabela 5.38, comprovam que as RNA que apresentam resultados ligeiramente melhores nas etapas de treino e teste, nem sempre são as que também apresentam melhores resultados na fase de predição. As redes “3I” e “2 8H sx”, segundo os resultados das etapas de treino e teste deveriam apresentar melhores resultados na fase de previsão a partir das séries sem a adição de dados, no entanto, não foi isso que aconteceu. Em termos de resultados propriamente ditos, há que destacar as redes “2E”, “4E”, “3I” e “1 9G8 sx”, sendo que as três últimas apresentaram os melhores resultados de toda a investigação global realizada. Os resultados da fase de treino e teste das redes “3I” e “2 8H sx”, produzidos a partir das séries com adição de dados, podem ser consultados no Anexo VIII. Na Figura 5.16 é possível comparar graficamente os valores reais com os valores previstos.

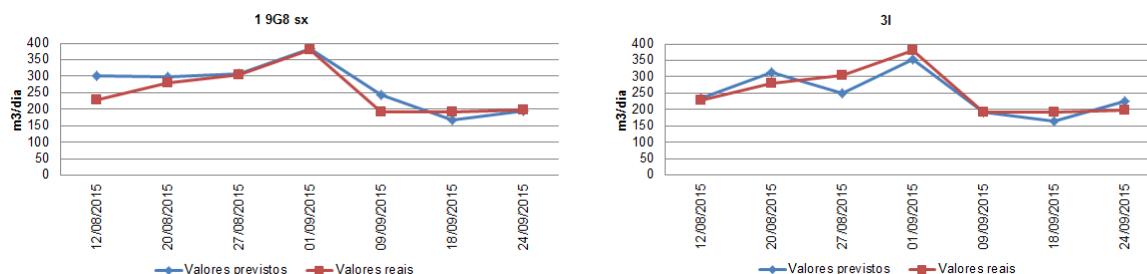


Figura 5.16 – Comparação entre os valores reais e os valores previstos de biogás produzido pelas redes “1 9G8 sx” e “3I” – ETAR de Sesimbra

Os resultados demonstram que apesar de o processo de digestão anaeróbia ser de grande variabilidade e deter muitos aspetos a controlar, as RNA detêm potencial para modelar de facto este processo. No período temporal em causa, houve RNA que conseguiram prever valores bastante próximos dos registados na realidade.

Tendo em conta que era possível executar a etapa de previsão com um maior número de casos destacados somente para previsão, fator que influencia os resultados finais, decidiu-se repetir a fase de previsão para as redes “2E” e “4E”, dado que foram as RNA produzidas a partir das séries sem adição de dados que melhores resultados apresentaram. No caso das redes “3I” e “1 9G8 sx”, geradas a partir das séries com adição de dados, também se queria executar uma etapa de previsão com um maior número de casos destacados somente para este estágio, no entanto não foi possível. A presunção assumida não se concretizou, visto se ter esgotado o tempo disponível para a realização da investigação e de em tempo útil não haver possibilidade de recolher mais dados. Na Tabela 5.39 são apresentados os resultados da nova fase de previsão realizada para as redes “2E” e “4E”, onde os casos destacados para previsão datam de 21 de Maio de 2015 a 24 de Setembro de 2015.

Tabela 5.39 – Resultados da nova fase de previsão para as redes “2E” e “4E”

Rede	Séries com ou sem adição de dados	Casos treino + teste	Casos previsão	RMS	R ²
2E	sem adição	52	15	63,67	0,23
4E	sem adição	34	15	94,14	0,06

O aumento do número de casos destacados para previsão gerou alterações nos resultados. Se com apenas sete casos, os resultados foram claramente superiores neste caso de estudo face aos resultados na ETAR do Seixal, o mesmo já não é tão evidente quando duplicamos os casos destacados somente destacados para esta etapa. O único parâmetro que não piorou foi o erro médio quadrático da rede “2E”, o qual melhorou, sendo ainda necessário realçar que com o aumento do número de casos para predição, a rede “2E” exibiu melhores resultados do que a rede “4E”. A Figura 5.17 exhibe a comparação gráfica entre os valores reais e os valores previstos.

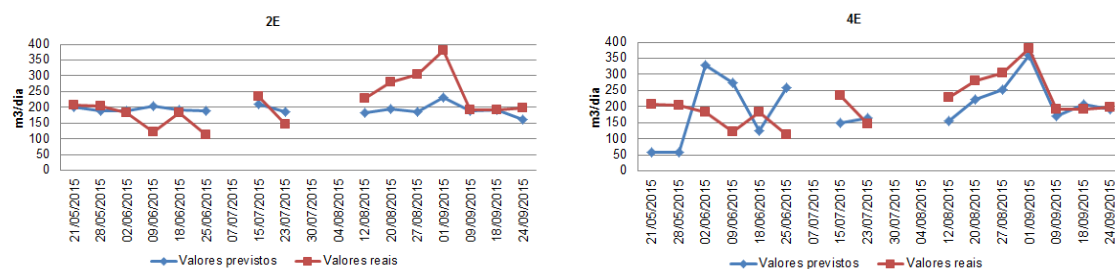


Figura 5.17 – Comparação entre os valores reais e os valores previstos de biogás produzido pelas redes “2E” e “4E” – ETAR de Sesimbra

As mudanças verificadas nos resultados advêm do facto de a digestão anaeróbia ser um processo bastante complexo e de grande variabilidade. O facto de as redes geradas estarem longe de serem precisas nas suas previsões, aliado ao aumento do número de casos de previsão, os quais representam um novo desafio do ponto de vista analítico, não possibilitou a manutenção dos bons resultados verificados na Tabela 5.38. Perante o novo período de casos de predição adicionados, as RNA não conseguiram responder com tão valores aproximados aos reais, como no período de 12 de Agosto a 24 de Setembro, pelo que os resultados pioraram. Todavia há que registar que a rede “2E” apresenta um resultado satisfatório, tendo conta o valor alcançado no erro médio quadrático.

Em suma, a partir dos resultados finais alcançados neste caso de estudo é possível retirar as seguintes ilações:

- Foram poucas as variáveis que apresentaram uma correlação moderada com a série de dados de caudal de biogás;
- Apesar de Sesimbra possuir uma população residente no Verão bastante superior em comparação aos restantes meses do ano, não se verificam alterações sazonais nos valores de caudal de biogás gerado;
- As séries de dados de frequência de amostragem diária apresentam bastantes valores fora dos intervalos de funcionamento ótimo de um processo de digestão anaeróbia;
- As séries de dados de frequência de amostragem semanal apresentam no geral, médias dentro dos padrões recomendados para o funcionamento do processo de digestão anaeróbia, contudo também possuem valores fora dos intervalos normais de laboração;
- Não foi possível destacar qual a variável com melhor desempenho, no entanto as variáveis “T reator”, “pH reator” e “MV reator” tiveram grande preponderância nas RNA desenvolvidas;
- As RNA criadas com base em séries com desfasamento de dados apresentam resultados piores em todas as fases de desenvolvimento, quando comparadas às RNA produzidas a partir de séries de dados sem desfasamento;

- Os resultados da fase de previsão das melhores redes, quando destacados casos apenas entre 12 de Agosto de 2015 e 24 de Setembro de 2015 foram bastante bons;
- Não é possível identificar nenhuma relação entre o número de casos de treino e teste utilizados e os resultados gerados na fase de previsão, no que diz respeito às RNA de dados de frequência de amostragem semanal, sem desfasamento;
- Os resultados globais da etapa de previsão demonstram que a variabilidade do próprio processo de digestão anaeróbia, é o maior problema para as capacidades modeladoras das RNA, dado que avaliar diferentes períodos temporais, pode significar diferentes performances das redes neuronais produzidas;
- Tendo em conta os resultados da fase de previsão, o número de casos disponíveis para treino e teste em cada rede e dado não ter sido possível executar esta etapa com um maior número de casos destacados nas RNA “3I” e “1 9G8 sx “, considera-se que a rede “2E” foi a que obteve um melhor resultado. A rede “2E” foi gerada a partir das três variáveis que obtiveram os valores mais altos de correlação com a série de dados do caudal de biogás, demonstrando que é preciso ter em conta os resultados deste cálculo.

5.3 Modelo Geral Comum aos Dois Casos de Estudo

Foi ainda efetuado mais um desenvolvimento que foi a criação de um modelo comum aos dois casos de estudo, para tal começou-se por comparar alguns valores do coeficiente de correlação, associados entre as variáveis utilizadas nos dois casos de estudo. As Tabelas 5.40 e 5.41 apresentam os valores de correlação mais elevados verificadas e as correlações registadas com a variável “Q biogás”, respetivamente.

Tabela 5.40 – Relação entre as correlações mais fortes verificadas nos dois casos de estudo

		ETAR Seixal	ETAR Sesimbra
TRH	Q lamas	-0,91	-0,20
AGV/TAL	AGV reator	0,78	0,58
LMV removida	LMV in	0,72	0,98
LMV out	Q lamas	0,71	0,73
LMV in	Q lamas	0,57	0,94
TAL reator	AGV/TAL	-0,47	-0,83
LMV removida	Q lamas	0,19	0,91

Do conjunto de correlações mais fortes verificadas (entre variáveis utilizadas) nos dois casos de estudo, quatro na ETAR do Seixal e cinco na ETAR de Sesimbra, apenas duas estão associadas a ambos os casos, e apenas a correlação entre “LMV out” e “Q lamas” apresenta um valor bastante idêntico. No geral das correlações verificadas entre variáveis, salientam-se bastantes diferenças.

Tabela 5.41 – Relação entre as correlações mais fortes verificadas com a variável “Q biogás”

		ETAR Seixal	ETAR Sesimbra
Q biogás	LMV removida	0,50	-0,01
Q biogás	pH in	0,46	0,09
Q biogás	AGV/TAL	-0,45	0,23
Q biogás	AGV reator	-0,44	-0,22
Q biogás	pH reator	0,43	0,64
Q biogás	T reator	0,42	0,45
Q biogás	MV in	0,39	0,28
Q biogás	LMV in	0,35	-0,16
Q biogás	MV reator	-0,09	-0,48

A nível das correlações apuradas das várias variáveis com o caudal de biogás gerado, a mais elevada é registada na ETAR de Sesimbra entre as variáveis “Q biogás” e “pH reator”, todavia na ETAR do Seixal o número de correlações moderadas é bastante maior. No global das correlações apuradas com o caudal de biogás, existem muitas diferenças entre casos de estudo, pelo que a tarefa de criar um modelo geral válido aos dois casos de estudo, no momento parecia improvável de acontecer. Apesar das más indicações retiradas da análise dos coeficientes de correlação de Pearson, decidiu-se utilizar as configurações de cada uma das melhores RNA geradas em cada ETAR e produzir duas novas redes neuronais com dados dos dois casos de estudo. As configurações usadas tiveram por base a rede “8F 2015”, a qual apresentou o melhor resultado na investigação realizada na ETAR do Seixal, e a rede “2E” responsável pelo melhor resultado registado na ETAR de Sesimbra.

A rede criada com a configuração de variáveis igual à rede “8F 2015” é designada de “Geral sx”, enquanto a rede gerada com a configuração equivalente à rede “2E” chama-se “Geral sb”. Na etapa de treino e teste foram utilizados todos os dados anteriormente usados para treinar e testar individualmente as redes “8F 2015” e “2E”, o que inclui no caso da rede “2E” os últimos dados adicionados às séries desta rede neuronal. Na fase de previsão, utilizaram-se todos os casos destacados na previsão individual das redes “8F 2015” e “2E”, o que inclui no caso da rede “2E” apenas os casos entre 12 de Agosto de 2015 e 24 de Setembro de 2015. A Tabela 5.42 exhibe os resultados da etapa de treino, teste e previsão das RNA “Geral sx” e “Geral sb”.

Tabela 5.42 – Resultados das fases de desenvolvimento das redes neuronais “Geral sx” e “Geral sb”

Rede	Variáveis de entrada	Impacto	Resultados - Treino e Teste			Resultados - Previsão		
Geral sx	TAL reator (mg/L)	32,19%	Nº casos	Melhor rede	GRNN	Nº casos	RMS	R ²
	AGV reator (mg/L)	22,08%	121	% más previsões (treino - 30% toler.)	8,25%	17	267,67	0,22
	pH reator	18,54%		% más previsões (teste - 30% toler.)	41,67%			
	TRH (dias)	17,70%		Erro médio quadrático (treino)	78,75			
	T reator (°C)	9,49%		Erro médio quadrático (teste)	122,35			
Geral sb	T reator (°C)	39,98%	Nº casos	Melhor rede	GRNN	Nº casos	RMS	R ²
	pH reator	31,41%	184	% más previsões (treino - 30% toler.)	47,62%	17	147,80	0,69
	MV reator (mg/L)	28,60%		% más previsões (teste - 30% toler.)	64,86%			
				Erro médio quadrático (treino)	184,24			
				Erro médio quadrático (teste)	220,70			

A rede neuronal “Geral sx” apesar de conter um menor número de casos disponíveis para a etapa de treino e teste obteve melhores resultados nestas fases, relativamente à rede neuronal “Geral sb”. Na Figura 5.18 pode observar-se a comparação entre os valores reais de biogás produzido e os valores previstos pelas duas RNA. As datas dos casos não são apresentadas, visto que os dados advêm de estudos diferentes e de um mesmo intervalo temporal, logo a datação de cada caso nesta situação não é um fator importante.

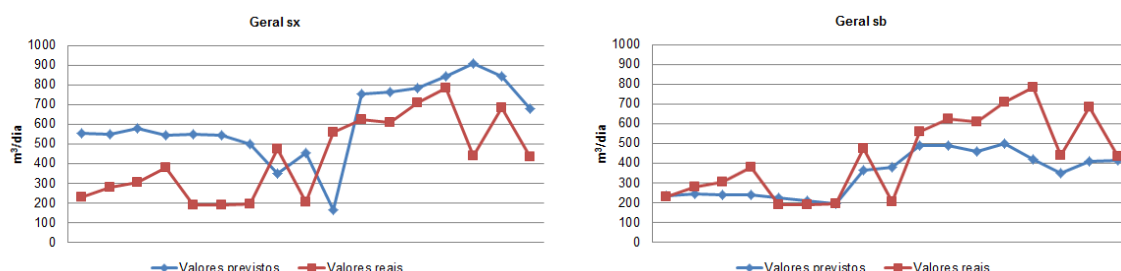


Figura 5.18 – Comparação dos valores reais e dos valores previstos de biogás produzido pelas redes “Geral sx” e “Geral sb” – ETAR de Sesimbra

Os resultados verificados na fase de previsão são surpreendentes, atendendo ao facto de que a rede “Geral sx” obteve resultados bastante superiores na fase de treino e teste, em comparação à rede “Geral sb”. Em outras situações, já se tinha verificado que por vezes RNA com piores resultados na fase de treino e teste, conseguiam obter melhores resultados na fase de previsão, no entanto os resultados eram muito mais semelhantes do que neste caso. Nada fazia prever que perante um resultado tão desnivelado na fase de treino e teste, iria-se também gerar um resultado desnivelado na etapa de previsão, mas em favor da rede que pior performance tinha obtido no treino e teste.

Este resultado pode significar hipoteticamente que o número de casos destacados para as previsões realizadas com as RNA de frequência de amostragem semanal é baixo, e sendo baixo os resultados dependem muito das características do período avaliado. Um número baixo

de casos destacados para previsão, perante um processo de grande variabilidade como a digestão anaeróbia, pode levar a que uma RNA de capacidade inferior, por uma questão de aleatoriedade consiga obter melhores resultados numa etapa de previsão. O período avaliado pode-se assemelhar mais às características dos casos treinados e testados de determinada RNA, adulterando desta forma os resultados finais. Outra explicação para os resultados registados é o facto de se ter misturado dados de dois casos de estudo com características e gamas de valores diferentes, sem se ter qualquer conhecimento dos efeitos que esta situação podia gerar. A mistura de dados pode ter levado a uma perda do significado das indicações que se costumavam retirar dos resultados das etapas de treino e teste, alterando a dinâmica relacional entre estas fases com a etapa de previsão.

Os resultados propriamente ditos não foram muito bons, tendo em conta que os erros médios quadráticos calculados foram elevados e os coeficientes de determinação deveriam ser mais elevados, para poderem ser considerados bons resultados. Porém os resultados foram melhores do que se esperava inicialmente, pelo que a hipótese de se construir um modelo geral a vários casos de estudo não parece de todo impossível.

A tentativa de criação do modelo geral marca o culminar da investigação executada. O trabalho realizado foi bastante extenso e abrangeu dois casos de estudo, por isso apresenta-se na Figura 5.19 em diagrama, um resumo de todo o estudo realizado.

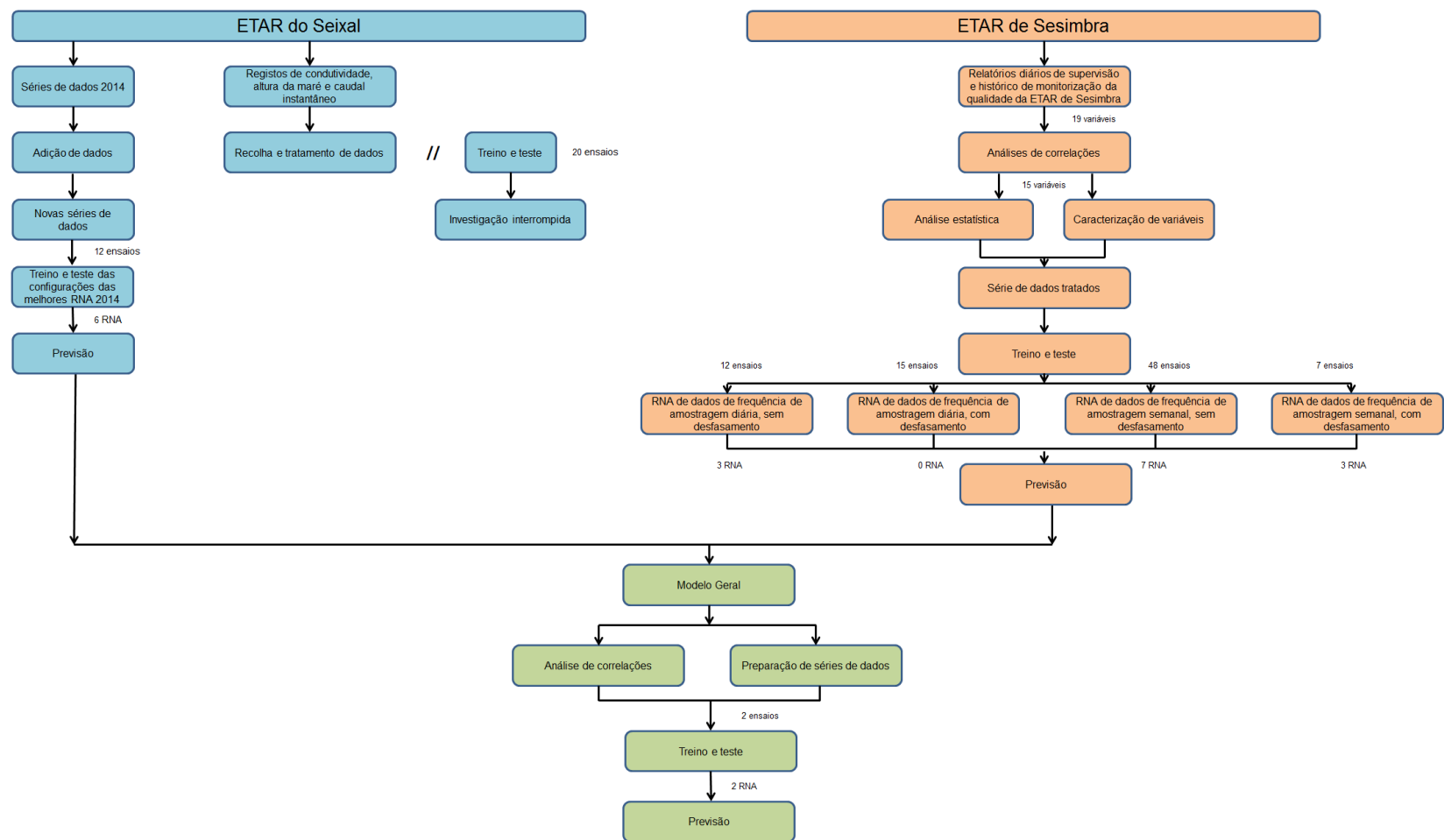


Figura 5.19 – Diagrama da investigação realizada

6. Conclusão e Sugestões para Desenvolvimentos Futuros

A investigação realizada foi bastante vasta, tendo em conta que o trabalho envolveu dois casos de estudo e uma quantidade assinalável de informação e de dados estatísticos. As principais conclusões que se podem retirar do trabalho efetuado, em cada caso de estudo e da criação de um modelo comum aos dois casos são apresentadas em seguida:

- ETAR do Seixal, primeira vertente do estudo:
 - Não foi possível concluir se o aumento de informação nas séries de dados foi ou não vantajoso. Os resultados alcançados face aos da investigação realizada no ano de 2014 foram inferiores para um parâmetro método estatístico e superiores para outro;
 - Os casos destacados para a fase de previsão datam de um período em que ocorreram alguns problemas operacionais na ETAR, o que destabilizou os valores das séries de dados afetas a esse intervalo de tempo e complicou a avaliação dos resultados. Os resultados registados na fase de treino e teste de cada RNA, levam a crer que neste caso de estudo a adição de dados às séries foi benéfica e teria produzido melhores resultados se os valores de biogás a prever, não derivassem de um período em que a digestão anaeróbia funcionou de modo conturbado.
- ETAR do Seixal, segunda vertente do estudo:
 - Os valores de condutividade registados não podem ser estimados com base em apenas dados de altura de maré e de caudal instantâneo;
 - A adição de dados às séries aumentou a variabilidade de valores a prever, acabando por piorar os resultados;
 - A incapacidade das RNA geradas, para prever valores de condutividade conduziu à suspensão da investigação na ETAR do Seixal;
 - O trabalho realizado foi infrutífero, não tendo sido possível fazer avanços significativos no conhecimento da relação entre os valores de condutividade e de biogás registados, ficando também por avaliar os valores de condutividade no caudal de lamas afluente à digestão anaeróbia.
- ETAR de Sesimbra:
 - A adição de dados às séries nem sempre produzirá melhores resultados;
 - Uma RNA que obteve piores resultados na fase de teste e treino, comparativamente a outra, pode na fase de previsão gerar melhores resultados;
 - As RNA criadas a partir de séries de desfasamento de dados, alcançaram no global resultados inferiores às RNA de séries sem desfasamento de dados;
 - Os resultados gerados com base nos valores de biogás previstos entre 12 de Agosto de 2015 a 24 de Setembro de 2015 foram bastante bons, porém quando se aumentou o número de casos a prever, a qualidade dos resultados diminuiu. A diferença de resultados entre períodos advém provavelmente do facto de que no período entre 12 de Agosto de 2015 a 24 de Setembro de 2015, o caudal de afluência de lamas foi consideravelmente superior. Esta maior afluência de lamas, aliada a bons resultados

em outros parâmetros, permitiu um melhor funcionamento do digestor anaeróbio, o que também se traduziu num melhor desempenho por parte das RNA na previsão de valores de biogás produzido.

➤ Modelo geral aos dois casos de estudo:

- A informação acerca dos efeitos que a junção de dados provenientes de dois ou mais casos de estudo pode ter nos resultados finais é praticamente inexistente. Contudo os resultados que as RNA geraram a partir da aglutinação de dados, não foi inferior aos resultados obtidos individualmente por conjunto de dados;
- Os resultados alcançados podem ser considerados como razoáveis, atendendo ao facto de que os dois casos de estudo possuem características e coeficientes de correlação entre variáveis bastante diferentes.

No global de toda a investigação realizada, salienta-se que o principal obstáculo à produção de melhores resultados por parte das RNA é a própria variabilidade do processo de digestão anaeróbia. As RNA analisam os dados do ponto de vista numérico, sem qualquer conhecimento adicional, todavia os parâmetros medidos e calculados afetos à digestão anaeróbia sofrem pequenas alterações todos os dias, o que dificulta a aprendizagem das RNA. Para complicar a situação, em qualquer um dos casos de estudo, existem muitos registos que comprovam a existência de problemas operacionais durante determinados períodos de tempo, o que pode levar a alterações na consistência das séries de dados. O facto de que as séries de dados utilizadas no presente trabalho possuem ainda poucos casos disponíveis para treino e teste das RNA é inegável, todavia a adição desmensurada de dados nem sempre é a melhor opção. A adição de dados só será vantajosa se esta contribuir para a consistência das séries de dados, ou seja, só existirão benefícios se a qualidade dos dados adicionados não for inferior à dos dados anteriormente em uso.

Perante situações como na presente investigação em que só foi possível destacar alguns casos para a etapa de previsão, os resultados finais alcançados derivam em parte das características de funcionamento do digestor anaeróbio, durante o período temporal de onde foram retirados estes mesmos casos. As características dos poucos casos de previsão podem-se assemelhar mais ou menos às características de determinada RNA. Este fator pode levar a que por vezes RNA com um potencial para predição de valores inferior, consigam obter melhores resultados.

Por fim considera-se que as RNA possuem uma enorme capacidade de modelação, podendo num futuro próximo vir a ser uma ferramenta importante na gestão de uma ETAR. No entanto, a aplicação desta tecnologia na modelação de um processo ainda apelidado por muitos investigadores de “*black-box*”, demorará algum tempo a produzir os resultados esperados, visto ser bastante difícil prever o funcionamento de um processo que não é inteiramente controlado.

A nível de sugestões para desenvolvimentos futuros, no seguimento das recomendações deixadas após a conclusão da investigação em 2014 na ETAR do Seixal, reafirma-se a importância do aumento da frequência de amostragem dos parâmetros qualitativos das lamas à entrada e no interior do digestor anaeróbio. A regularização dos intervalos temporais entre amostras também é um aspeto que deve ser normalizado. Estas medidas são fundamentais para se obter melhores resultados quando se recorre ao desenvolvimento de RNA. O estudo e integração da microbiologia como um dos aspetos a ter em conta na modelação do funcionamento da digestão anaeróbia é outra das sugestões que se repetem. As comunidades microbiológicas apresentam características distintas de caso de estudo para caso de estudo, pelo que seria interessante estudar os efeitos que estas diferenças provocam.

No caso específico da ETAR do Seixal, tendo em conta os problemas gerados pelos altos valores de condutividade existentes no digestor anaeróbio, recomenda-se que os problemas relacionados com as intrusões salinas sejam resolvidos a montante, em vez de se tentar remendar o problema na etapa de digestão anaeróbia. É com agrado que se constata que algumas medidas já estão a ser implementadas a montante, tendo-se inclusive verificado melhorias na variação de condutividade no caudal afluente à ETAR. Apesar de se compreender que não é um trabalho que se realiza de forma rápida e que tal decisão acarreta maiores custos financeiros a curto prazo, acredita-se que esta estratégia a longo prazo será vantajosa a nível técnico e económico. No futuro seria também interessante realizar uma investigação, tendo por alvo medir os valores de condutividade diretamente no caudal de lamas afluente ao digestor, de forma a relacionar este parâmetro com os valores de biogás, algo que não se conseguiu durante este trabalho.

Sugere-se que no futuro seja tida em consideração a hipótese de o intervalo temporal entre investigações num mesmo caso de estudo, ser superior a um ano. Apesar de se considerar o acréscimo de dados vantajoso, dificilmente é possível adicionar todos os dados disponibilizados num período de um ano, tendo em conta que os valores fora dos padrões normais da base de dados não devem ser considerados para o treinamento das RNA. Esta situação acaba por limitar o número de dados a adicionar às séries. Considera-se também interessante a possibilidade de se criar uma variável que através de um código simples numérico (binário por exemplo), conceda informação à RNA acerca do regime de funcionamento em que o digestor anaeróbio se encontra. Esta variável permitiria à RNA saber se os valores registados dos vários parâmetros, advêm de um funcionamento normal do processo de digestão anaeróbia ou se estão a ser afetados por problemas operacionais. Acredita-se que seria uma informação útil de ser concedida à rede neuronal e que talvez tivesse um impacto positivo nos resultados.

Um dos aspetos relevantes que se apurou durante a investigação é o facto que o número de casos para previsão que se tem utilizado nas RNA, criadas a partir das séries com dados de

frequência de amostragem semanal parece ser pouco significativo, devendo ser mais elevado. O trabalho realizado leva a crer que provavelmente em alguns casos, é necessário um espaço temporal maior entre investigações de forma a ser possível a adição de um número considerável de dados, ao mesmo tempo que também é possível destacar um maior número de casos somente para a fase de previsão. Caso esta sugestão seja seguida, recomenda-se que durante o tempo de interregno sejam executados trabalhos com o objetivo de desenvolver e testar modelos que englobem dados relativos a processos de digestão anaeróbia, localizados em diferentes casos de estudo. Neste sentido, além de se executar uma fase de previsão para valores de ambos os casos de estudo, deve-se também realizar a previsão para valores de cada caso de estudo em separado, algo que no presente trabalho não se pode executar em tempo oportuno, mas que no futuro seria interessante analisar.

7. Referências Bibliográficas

- Abbasi, T., S.M. Tauseef e S.A. Abbasi. 2012. "A Brief History of Anaerobic Digestion and Biogas" Pp. 11-23 in Biogas Energy. EUA: Springer Science+Business Media.
- Adgar, A., C. S. Cox e T. J. Bo. 2000. "Performance Improvements At Surface Water Treatment Works Using Ann-Based." Institution of Chemical Engineers 78(October):1026–1039.
- AdP – Águas de Portugal. 2015. Site consultado em Maio de 2015 (<http://www.adp.pt/>).
- Ahlgren, Per, Bo Jarneving e Ronald Rousseau. 2003. "Requirements for a Cocitation Similarity Measure , with Special Reference to Pearson's Correlation Coefficient." Journal of the American Society for Information Science and Technology 54(6):550–560.
- Apa – Agência Portuguesa do Ambiente. 2015. "Águas costeiras e de transição". Site consultado em Julho de 2015 (<http://www.apambiente.pt/?ref=16&subref=7&sub2ref=875&sub3ref=876>).
- Appels, Lise, Jan Baeyens, Jan Degreève e Raf Dewil. 2008. "Principles and Potential of the Anaerobic Digestion of Waste-Activated Sludge." Progress in Energy and Combustion Science 34(6):755–81.
- Batstone, D. J., J.Keller, I. Angelidaki, S.V. Kalyuzhnyi, S.G. Pavlostathis, A.Rozzi, W.T.M. Sanders, H. Siegrist e V.A. Vavilin. 2002. "The IWA Anaerobic Digestion Model No 1 (ADM1)." Water Science and Technology 45(10):65–73.
- Blumensaat, F e J. Keller. 2005. "Modelling of Two-Stage Anaerobic Digestion Using the IWA Anaerobic Digestion Model No. 1 (ADM1)." Water Research 39(1):171–83.
- Bose, N. K. e P. Liang. 1996. Neural Network Fundamentals with Graphs, Algorithms, and Applications. EUA: McGraw-Hill.
- Chen, Ye, Jay J. Cheng e Kurt S. Creamer. 2008. "Inhibition of Anaerobic Digestion Process: A Review." Bioresource Technology 99(10):4044–64.
- Colleran, E., S. Finnegan, e P. Lens. 1995. "Anaerobic Treatment of Sulphate-Containing Waste Streams." Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 67(1):29–46.
- Colleran, E., S. Pender, U. Philpott, V. O'Flaherty e B. Leahy. 1998. "Full-Scale and Laboratory-Scale Anaerobic Treatment of Citric Acid Production Wastewater." Biodegradation 9(3-4):233–45.

Decreto-Lei nº 152/97 de 19 de Junho. Diário da República nº 139 – 19-6-1997 – 1ª Série-A. Ministério do Ambiente.

Decreto-Lei nº 348/98 de 9 de Novembro. Diário da República nº 259 – 9-11-1998 – 1ª Série-A. Ministério do Ambiente.

Decreto-Lei nº 149/2004 de 22 de Junho. Diário da República nº 145 – 22 de Junho de 2014 – 1ª Série-A. Ministério das Cidades, Ordenamento do Território e Ambiente.

Decreto-Lei nº 198/2008 de 8 de Outubro. Diário da República nº 195 – 8 de Outubro de 2008 – 1ª série. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional.

Decreto-Lei nº 133/2015 de 13 de Julho. Diário da República nº 134 – 13 de Julho de 2015 – 1ª série. Ministério do Ambiente, Ordenamento do Território e Energia.

Degrèmont, Ondeo. 2002. Water Treatment Handbook, 6ª Edição. França: Lavoisier.

Derbal, K., M. Bencheikh-lehocine, F. Cecchi, a. H. Meniai e P. Pavan. 2009. “Application of the IWA ADM1 Model to Simulate Anaerobic Co-Digestion of Organic Waste with Waste Activated Sludge in Mesophilic Condition.” *Bioresource Technology* 100(4):1539–43.

Ding, Shifei, Hui Li, Chunyang Su, Junzhao Yu e Fengxiang Jin. 2011. “Evolutionary Artificial Neural Networks: A Review.” *Artificial Intelligence Review* 251–60.

Diretiva 91/271/CEE de 21 de Maio. Jornal Oficial das Comunidade Europeias nº L 135/40. Conselho das Comunidades Europeias.

Donoso-Bravo, a., C. Retamal, M. Carballa, G. Ruiz-Filippi e R. Chamy. 2009. “Influence of Temperature on the Hydrolysis, Acidogenesis and Methanogenesis in Mesophilic Anaerobic Digestion: Parameter Identification and Modeling Application.” *Water Science and Technology* 60(1):9–17.

Donoso-Bravo, Andres., Johan Mailier, Cristina Martin, Jorge Rodríguez, César Arturo Aceves-Lara e Alain Vande Wouwer. 2011. “Model Selection, Identification and Validation in Anaerobic Digestion: A Review.” *Water Research* 45(17):5347–64.

Feijoo, Gumersindo, Manuel Soto, Ramón Méndez e Juan M. Lema. 1995. “Sodium Inhibition in the Anaerobic Digestion Process: Antagonism and Adaptation Phenomena.” *Enzyme and Microbial Technology* 17(2):180–88.

FERROVIAL-AGROMAN/CADAGUA. 2007. Memória Descritiva e Justificativa do Projeto de Execução da ETAR do Seixal, Vol. 1 – Processo de Tratamento.

- Güçlü, Dünyamin, Nihat Yılmaz, e Umay G. Ozkan-Yucel. 2011. "Application of Neural Network Prediction Model to Full-Scale Anaerobic Sludge Digestion." *Journal of Chemical Technology & Biotechnology* 86(5):691–98.
- Haykin, Simon. 1999. *Neural Networks – A Comprehensive Foundation*, 2ª Edição. EUA: Pretice-Hall.
- Henze, Mogens e Yves Comeau. 2008. "Wastewater Characterization." *Biological Wastewater Treatment: Principles Modelling and Design*. 33–52.
- Holm-Nielsen, J. B., T. Al Seadi, e P. Oleskowicz-Popiel. 2009. "The Future of Anaerobic Digestion and Biogas Utilization." *Bioresource Technology* 100(22):5478–84.
- Holubar, Peter., Loredana Zani, Michael Hager, Walter Fröschl, Zorana Radak e Rudolf Braun. 2002. "Advanced Controlling of Anaerobic Digestion by Means of Hierarchical Neural Networks." *Water Research* 36(10):2582–88.
- Instituto Hidrográfico. 2014. *Tabelas de Maré 2015, Volume I – Portugal, Capítulo I – Generalidades*. Disponível em <http://www.hidrografico.pt/download-tabelas-mare.php>.
- Jäger-Waldau, Arnulf, Márta Szabó, Nicolae Scarlat e Fabio Monforti-Ferrario. 2011. "Renewable Electricity in Europe." *Renewable and Sustainable Energy Reviews* 15(8):3703–16.
- Jeison, D., A. Del Rio e J. B. Van Lier. 2008. "Impact of High Saline Wastewaters on Anaerobic Granular Sludge Functionalities." *Water Science and Technology* 57(6):815–19.
- Jenicek, P., C. A. Celis, L. Krayzelova, N. Anferova e D. Pokorna. 2014. "Improving Products of Anaerobic Sludge Digestion by Microaeration." *Water Science & Technology* 69(4):803.
- Jingura, Raphael M. e Rutendo Matengaifa. 2009. "Optimization of Biogas Production by Anaerobic Digestion for Sustainable Energy Development in Zimbabwe." *Renewable and Sustainable Energy Reviews* 13(5):1116–20.
- Kasabov, Nikola K. 1998. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, 2ª Edição. EUA: Massachusetts Intitute of Technology.
- Kythreotou, Nicoletta, Georgios Florides e Savvas A. Tassou. 2014. "A Review of Simple to Scientific Models for Anaerobic Digestion." *Renewable Energy* 71:701–14.
- Lapa, Nuno. 2014. *Digestão Anaeróbia*, documento de apoio à cadeira de Bioenergia, 2014/2015, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

- Lauwers, Joost, Lise Appels, Ian P. Thompson, Jan Degreë, Jan F. Van Impe e Raf Dewil. 2013. "Mathematical Modelling of Anaerobic Digestion of Biomass and Waste: Power and Limitations." *Progress in Energy and Combustion Science* 39(4):383–402.
- Lee, Myung Yeol, Chang Won Suh, Yong Tae Ahn, e Hang Sik Shin. 2009. "Variation of ADM1 by Using Temperature-Phased Anaerobic Digestion (TPAD) Operation." *Bioresource Technology* 100(11):2816–22.
- Lefebvre, Olivier e René Moletta. 2006. "Treatment of Organic Pollution in Industrial Saline Wastewater: A Literature Review." *Water Research* 40(20):3671–82.
- Lin Chen, Jian, Raphael Ortiz, Terry W. J. Steele e David C. Stuckey. 2014. "Toxicants Inhibiting Anaerobic Digestion : A Review." *Biotechnology Advances* 32(8):1523–34.
- Lindmark, Johan, Eva Thorin, Rebei Bel Fdhila e Erik Dahlquist. 2014. "Effects of Mixing on the Result of Anaerobic Digestion: Review." *Renewable and Sustainable Energy Reviews*.
- Lingireddy, Srinivasa e Gail M. Brion. 2005. *Artificial Neural Networks in Water Supply Engineering*. EUA: American Society of Civil Engineers.
- Malina, Joseph F. e Frederick G. Pohland. 1992. *Design of Anaerobic Processes for the Treatment of Industrial and Municipal Wastes, Volume 7*. EUA: Technomic Publishing Company, Inc.
- Martins, Fernando. 1997. "Aplicação de Redes Neurais Artificiais em Simulação e Controlo de Processos Químicos", Tese de Doutoramento. Portugal: Faculdade de Engenharia da Universidade do Porto.
- Metcalf & Eddy. 2003. *Wastewater Engineering – Treatment and Reuse*, 4ª Edição. EUA: McGraw-Hill.
- Murrayt, Patti e Stephen H. Zinder. 1985. "Nutritional Requirements of Methanosarcina Strain TM-1." *Analysis* 50(1):49–55.
- Nikpey, H., M. Assadi e P. Breuhaus. 2013. "Development of an Optimized Artificial Neural Network Model for Combined Heat and Power Micro Gas Turbines." *Applied Energy* 108:137–48.
- O'Flaherty, Vincent, Thérèse Mahony, Ronan O'Kennedy e Emer Colleran. 1998. "Effect of pH on Growth Kinetics and Sulphide Toxicity Thresholds of a Range of Methanogenic, Syntrophic and Sulphate-Reducing Bacteria." *Process Biochemistry* 33(5):555–69.

- Oh, Giyang, Lei Zhang e Deokjin Jahng. 2008. "Osmoprotectants enhance methane production from the anaerobic digestion of food wastes containing a high content of salt." *Journal of Chemical Technology and Biotechnology* 83:1204 – 1210.
- Palisade. 2013. *NeuralTools – Neural Network Add-In for Microsoft Excel®*, 6ª Versão. EUA: Palisade Corporation.
- Palisade. 2015. "NeuralTools." Site consultado em Junho de 2015 (<http://www.palisade.com/neuralTools/>).
- Palmeiro-Sánchez, T., A. Val del Río, A. Mosquera-Corral, J. L. Campos e R. Méndez. 2013. "Comparison of the Anaerobic Digestion of Activated and Aerobic Granular Sludges under Brackish Conditions." *Chemical Engineering Journal* 231:449–54.
- PEAASAR II – Plano Estratégico de Abastecimento de Água e de Saneamento de Águas Residuais. 2007. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional.
- PENSAAR 2020 – Uma nova Estratégia para o Setor de Abastecimento de Água e Saneamento de Águas Residuais. 2015. Ministério do Ambiente, Ordenamento do Território e Energia.
- Pilson, Michael E. Q. 2012. *An Introduction to the Chemistry of the Sea*, 2ª Edição. EUA: Cambridge University Press, New York.
- Pinto, Pedro. 2014. "Utilização de redes neuronais artificiais como ferramenta de gestão de processos de digestão anaeróbia. Caso de estudo – ETAR do Seixal", Dissertação para obtenção do Grau Mestre em Engenharia do Ambiente – Perfil de Engenharia Sanitária. Portugal: FCT – UNL.
- Puig, S., M. Serra, M. Coma, M. D. Balaguer e J. Colprim. 2011. "Simultaneous Domestic Wastewater Treatment and Renewable Energy Production Using Microbial Fuel Cells (MFCs)." *Water Science and Technology* 64(4):904–9.
- Puth, Marie Therese, Markus Neuhauser e Graeme D. Ruxton. 2014. "Effective Use of Pearson's Product-Moment Correlation Coefficient." *Animal Behaviour* 93:183–89.
- Qasim, Syed R. 1999. *Wastewater Treatment Plants – Planning, Design, and Operation*, 2ª Edição. EUA: CRC Press LLC.

- Rajeshwari, K. V., M. Balakrishnan, A. Kansal, K. Lata e V. V. Kishore. 2000. "State-of-the-Art of Anaerobic Digestion Technology for Industrial Wastewater Treatment." *Renewable and Sustainable Energy Reviews* 4(2):135–56.
- Rao, Valluru e Hayagriva Rao. 1995. *C++ Neural Networks and Fuzzy Logic*, 2ª edição. EUA: MIS:Press.
- Rojas, Raúl. 1996. *Neural Networks – A Systematic Introduction*. Alemanha: Springer-Verlag.
- Schmidt, J. E., a. J. L. Macario, B. K. Ahring e E. C. De Macario. 1992. "Effect of Magnesium on Methanogenic Subpopulations in a Thermophilic Acetate-Degrading Granular Consortium." *Applied and Environmental Microbiology* 58(3):862–68.
- Serrano, Cristina. 2014. "Impactes da Presença de Água do Estuário do Rio Tejo em ETAR. Caso de Estudo – ETAR do Seixal", Dissertação para obtenção do Grau Mestre em Engenharia do Ambiente, perfil de Engenharia Sanitária. Portugal: FCT – UNL.
- Shan, Yichu, Ruihuan Zhao, Guowang Xu, H. Liebich e Yukui Zhang. 2002. "Application of Probabilistic Neural Network in the Clinical Diagnosis of Cancers Based on Clinical Chemistry Data." *Analytica Chimica Acta* 471(1):77–86.
- Shiffman, Daniel. 2012. *The Nature of Code*. EUA: Magic Book Project. Site consultado em Agosto de 2015 (<http://natureofcode.com/book/>).
- SIMARSUL - Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal. 2007. *Manual de Operação da ETAR de Sesimbra*. Portugal: SIMARSUL.
- SIMARSUL – Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal. 2015. Site consultado em Maio de 2015 (<http://www.simarsul.pt/>).
- SIMRIA – Sistema Multimunicipal de Saneamento da Ria de Aveiro, S.A. 2015. Site consultado em Maio de 2015 (<http://www.simria.pt/>).
- Specht, Donald F. 1990. "Probabilistic Neural Networks." *Neural Networks* 3(1):109–18.
- Taylor, J. G. 1996. *Neural Networks and Their Applications*. Inglaterra: John Wiley & Sons.
- Turovskiy, Izrail S. e P.K. Mathai. 2006. *Wastewater Sudge Processing*. EUA: John Wiley & Sons.
- UNEP GEMS/Water Programme – United Nations Environment Programme Global Environment Monitoring System (GEMS)/ Water Programme. 2008. *Water Quality for*

Ecosystem and Human Health, 2ª Edição. Canadá: United Nations Environment Programme Global Monitoring System/Water Programme.

UNESCO – United Nations Educational, Scientific and Cultural Organization. 1985. The International System of Units (SI) in Oceanography. França: United Nations Educational, Scientific and Cultural Organization.

Vallero, M. V. G., L. W. Hulshoff Pol, G. Lettinga e P. N. L. Lens. 2003. “Effect of NaCl on Thermophilic (55°C) Methanol Degradation in Sulfate Reducing Granular Sludge Reactors.” *Water Research* 37(10):2269–80.

Van Langerak, E. P., G. Gonzalez-Gil, A. Van Aelst, J. B. Van Lier, H. V. M. Hamelers e G. Lettinga. 1998. “Effects of High Calcium Concentrations on the Development of Methanogenic Sludge in Upflow Anaerobic Sludge Bed (UASB) Reactors.” *Water Research* 32(4):1255–63.

Vemuri, Rao V. 1992. *Artificial Neural Networks: Concepts and Control Applications*. EUA: Institute of Electrical and Electronics Engineers Computer Society Press.

Vesilind, Aarne P. 2003. *Wastewater Treatment Plant Design*. Inglaterra: Water Environment Federation.

Von Sperling, Marcos. 2007. *Wastewater Characteristics, Treatment and Disposal, Biological Wastewater Treatment Series, Volume 1*. Inglaterra: IWA Publishing.

Worden, K., C.X. Wong, U. Parlitz, A. Hornstein, D. Engster, T. Tjahjowidodo, F. Al-Bender, D. Rizos e S. D. Fassois. 2007. “Identification of Pre-Sliding and Sliding Friction Dynamics: Grey Box and Black-Box Models.” *Mechanical Systems and Signal Processing* 21(1):514–34.

Yu, Liang, Pc Wensel, Jingwei Ma e Shulin Chen. 2013. “Mathematical Modeling in Anaerobic Digestion (AD).” *Journal of Bioremediation & Biodegradation* S4(003):12.

Zhang, Jingxin, Yaobin Zhang e Xie Quan. 2012. “Electricity Assisted Anaerobic Treatment of Salinity Wastewater and Its Effects on Microbial Communities.” *Water Research* 46(11):3535–43.

Zhang, Xuedong, Rui B. Ferreira, Henri Spanjers e Jules B. Van Lier. 2013. “Phosphatase Activity and Specific Methanogenic Activity in an Anaerobic Reactor Treating Sludge from a Brackish Recirculation Aquaculture System.” *Water Science and Technology* 68(2):387–93.

Anexos

Anexo I - Relatório detalhado da fase de treino e teste

Tabela I.1 – Exemplo de um relatório detalhado gerado após o término da fase de treino e teste (continua)

NeuralTools: Neural Net Training and Auto-Testing Performed By: Utilizador Date: 18 de junho de 2015 07:39:57 Data Set: sesimbra 2A Net: Net Trained on sesimbra 2A	
Summary	
Net Information	
Name	Net Trained on sesimbra 2A
Configurations Included in Search	GRNN, MLFN 2 to 15 nodes
Best Configuration	GRNN Numeric Predictor
Location	This Workbook
Independent Category Variables	0
Independent Numeric Variables	2 (Q lamas (m3/dia), T reator (°C))
Dependent Variable	Numeric Var. (Q biogás (m3/dia))
Training	
Number of Cases	360
Training Time	00:00:01
Number of Trials	52
Reason Stopped	Auto-Stopped
% Bad Predictions (30% Tolerance)	56,1111%
Root Mean Square Error	177,26
Mean Absolute Error	93,47
Std. Deviation of Abs. Error	150,61
Testing	
Number of Cases	90
% Bad Predictions (30% Tolerance)	53,3333%
Root Mean Square Error	125,39
Mean Absolute Error	82,80
Std. Deviation of Abs. Error	94,16
Data Set	
Name	sesimbra 2A
Number of Rows	870
Manual Case Tags	NO
Variable Impact Analysis	
T reator (°C)	50,7851%
Q lamas (m3/dia)	49,2149%

Tabela I.1 – (continuação)

Best Net Search			
	RMS Error	Training Time	Reason Training Stopped
GRNN	125,39	00:00:01	Auto-Stopped
MLFN 2 Nodes	135,89	00:30:00	Auto-Stopped
MLFN 3 Nodes	795,93	00:30:00	Auto-Stopped
MLFN 4 Nodes	126,63	00:30:00	Auto-Stopped
MLFN 5 Nodes	126,68	00:30:00	Auto-Stopped
MLFN 6 Nodes	896,87	00:30:00	Auto-Stopped
MLFN 7 Nodes	686,96	00:30:00	Auto-Stopped
MLFN 8 Nodes	157,13	00:30:00	Auto-Stopped
MLFN 9 Nodes	191,57	00:30:00	Auto-Stopped
MLFN 10 Nodes	513,13	00:30:00	Auto-Stopped
MLFN 11 Nodes	707,25	00:30:00	Auto-Stopped
MLFN 12 Nodes	328,99	00:30:00	Auto-Stopped
MLFN 13 Nodes	1136,83	00:30:00	Auto-Stopped
MLFN 14 Nodes	394,59	00:30:00	Auto-Stopped
MLFN 15 Nodes	160,91	00:30:00	Auto-Stopped

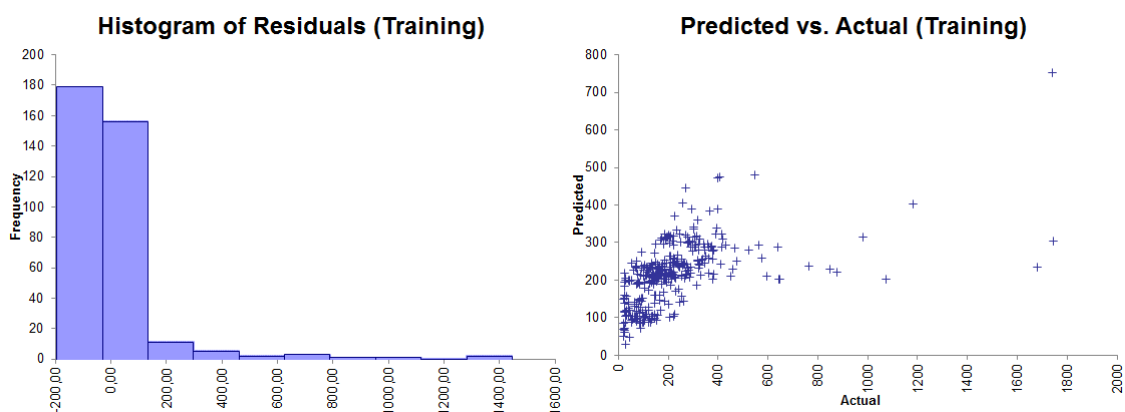


Figura I.1 – Relação entre os valores reais, previstos e residuais (continua)

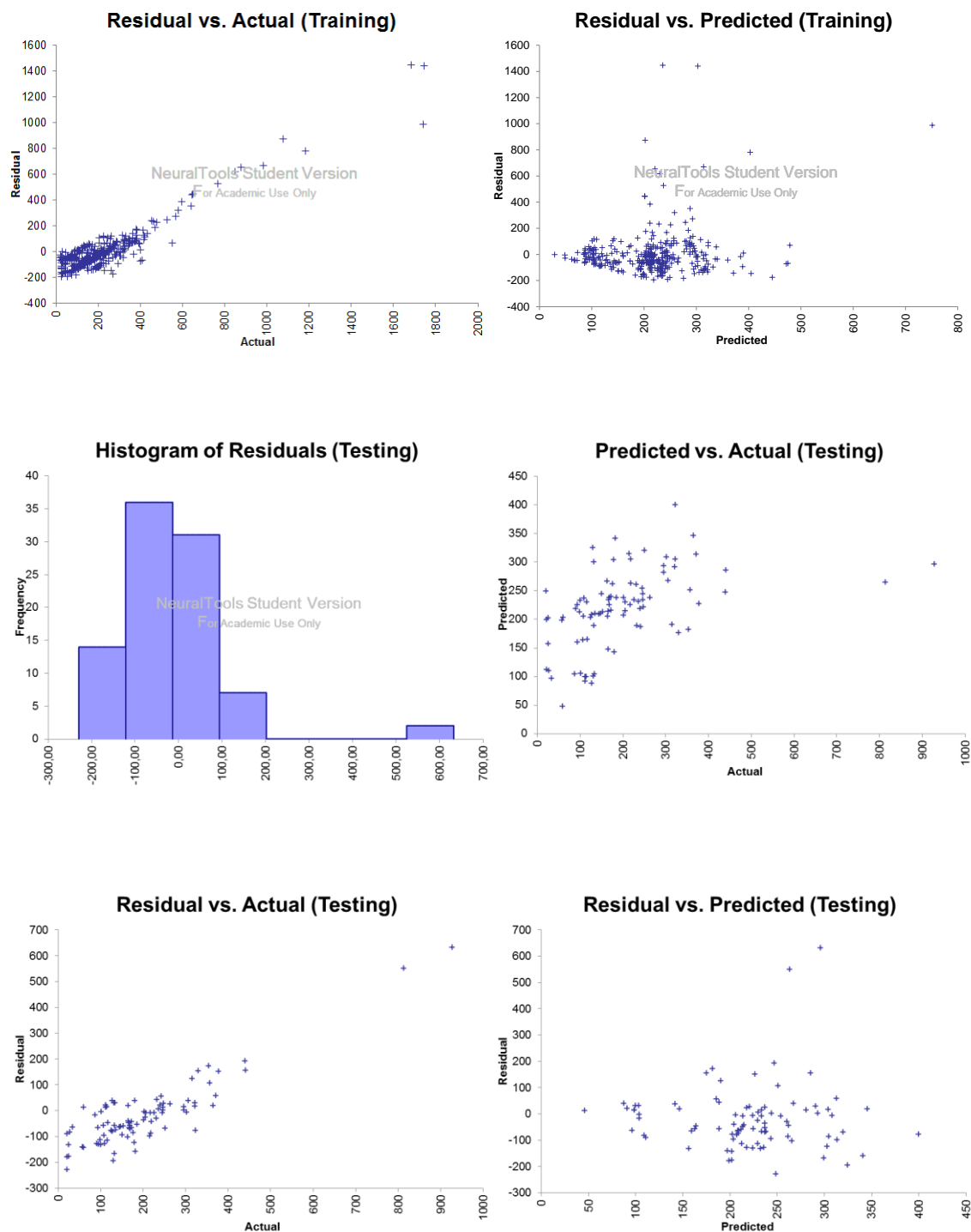


Figura I.1 – (continuação)

Anexo II - Localização dos pontos de amostragem e medição das variáveis na ETAR do Seixal

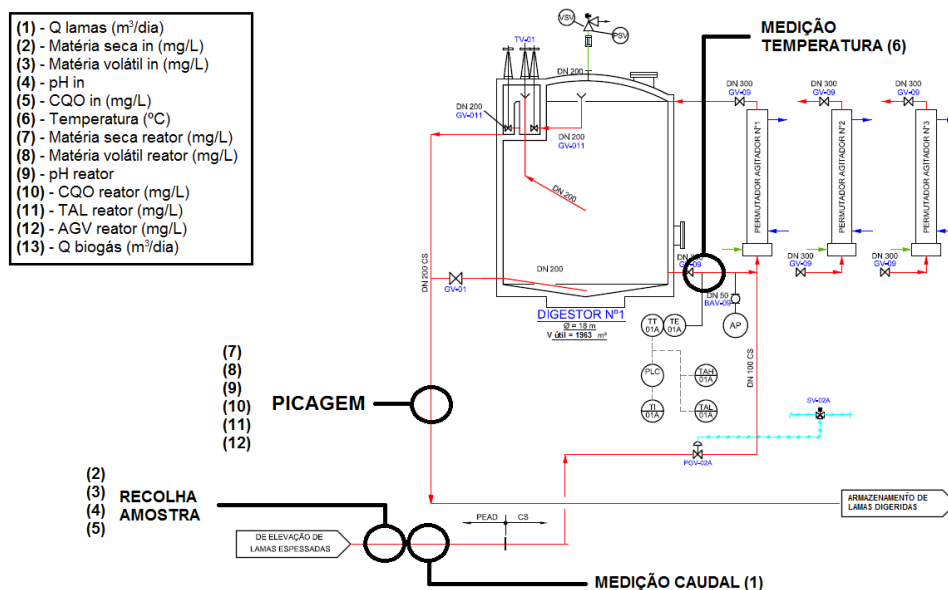


Figura II.1 – Localização das medições efetuadas e locais de amostragem de alguns parâmetros afetos à digestão anaeróbia – ETAR do Seixal

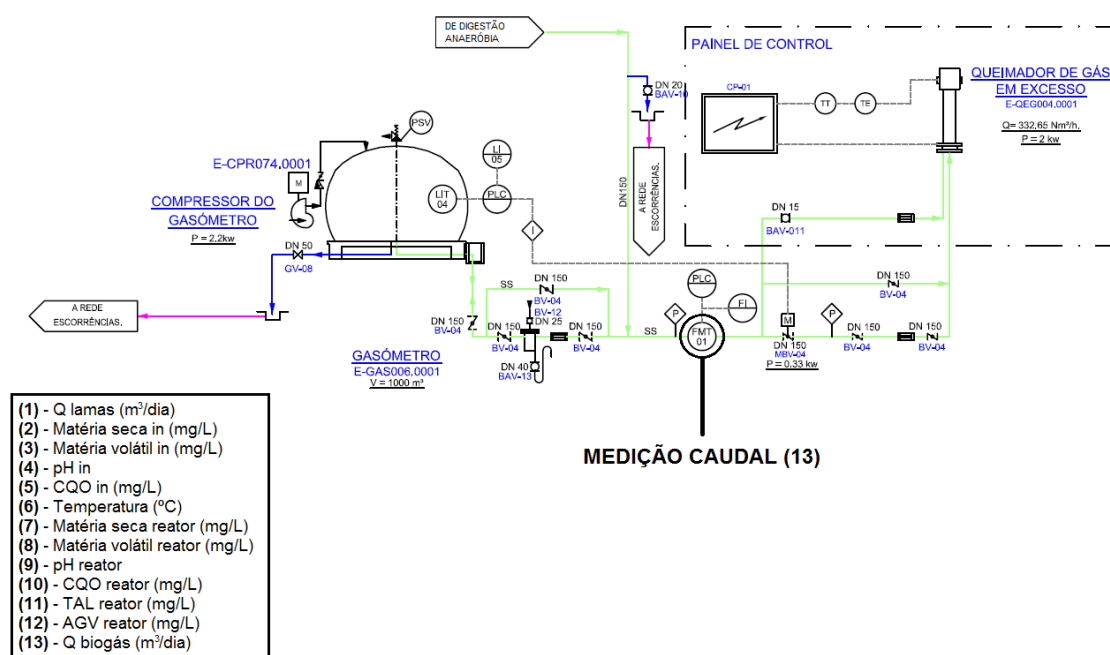


Figura II.2 – Local onde é efetuada a medição afeta ao caudal de biogás – ETAR do Seixal

Anexo III – Representação gráfica do registo de dados das variáveis utilizadas na primeira vertente da investigação realizada na ETAR do Seixal

As Figuras apresentadas indicam que no geral existe bastante variabilidade nos valores registados de cada variável, os quais fogem por vezes dos parâmetros normais de funcionamento, tal como se indicou na investigação realizada neste caso de estudo no ano de 2014.

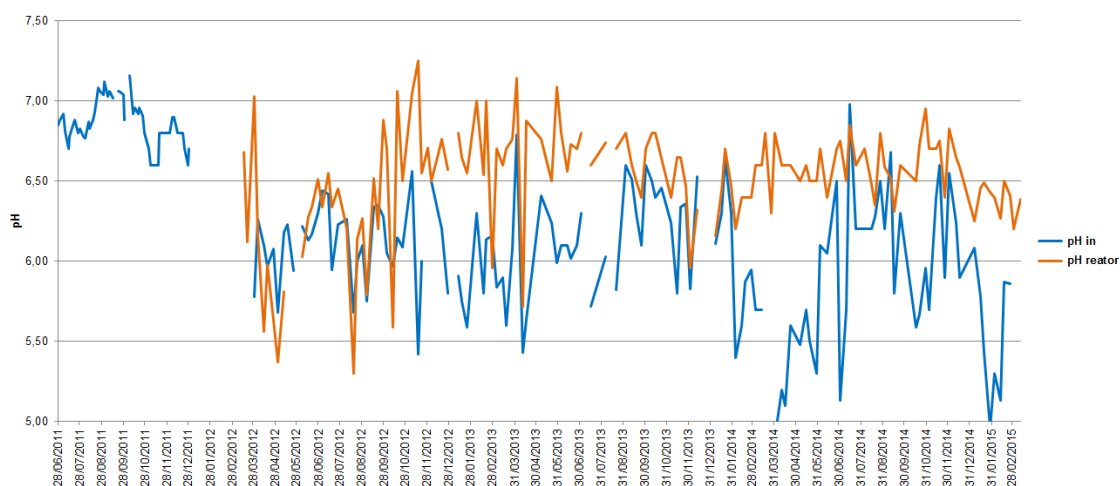


Figura III.1 – Comparação entre os dados das variáveis “pH in” e “pH reator” – ETAR do Seixal

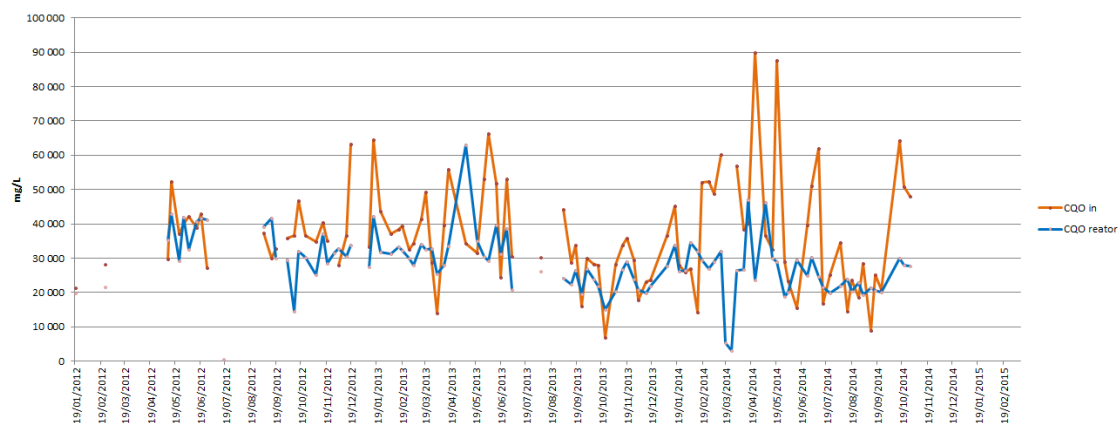


Figura III.2 – Comparação entre os dados das variáveis “CQO in” e “CQO reator” – ETAR do Seixal

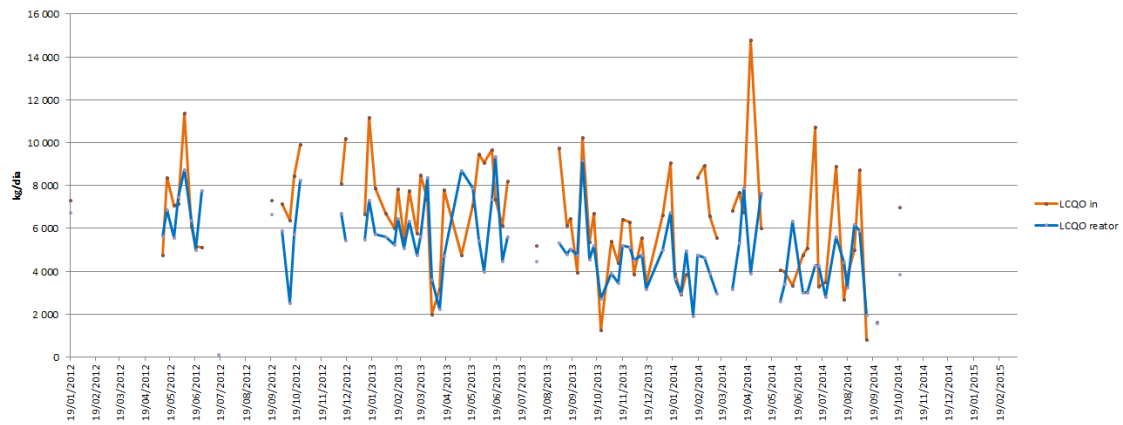


Figura III.3 - Comparação entre os dados das variáveis “LCQO in” e “LCQO reator” – ETAR do Seixal

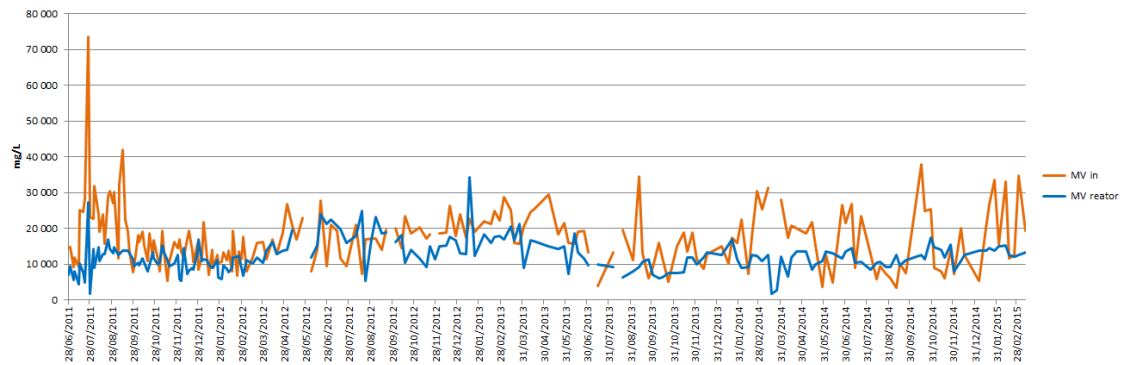


Figura III.4 - Comparação entre os dados das variáveis “MV in” e “MV reator” – ETAR do Seixal

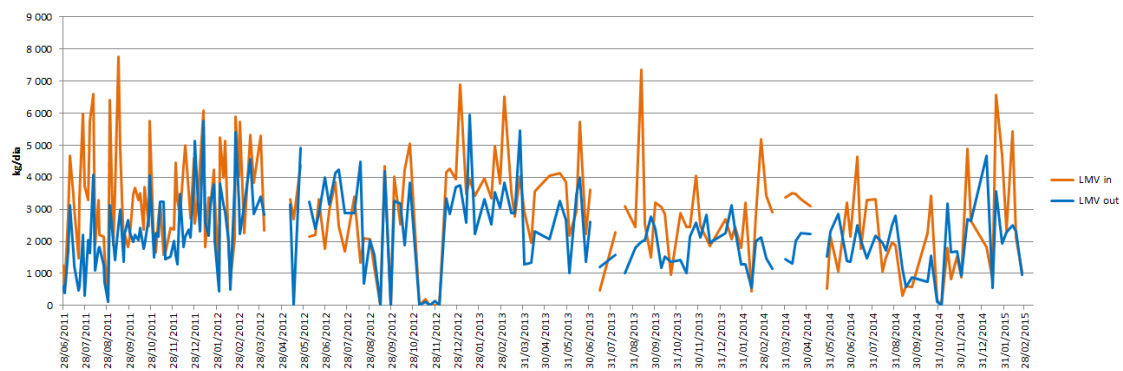


Figura III.5 - Comparação entre os dados das variáveis “LMV in” e “LMV out” – ETAR do Seixal

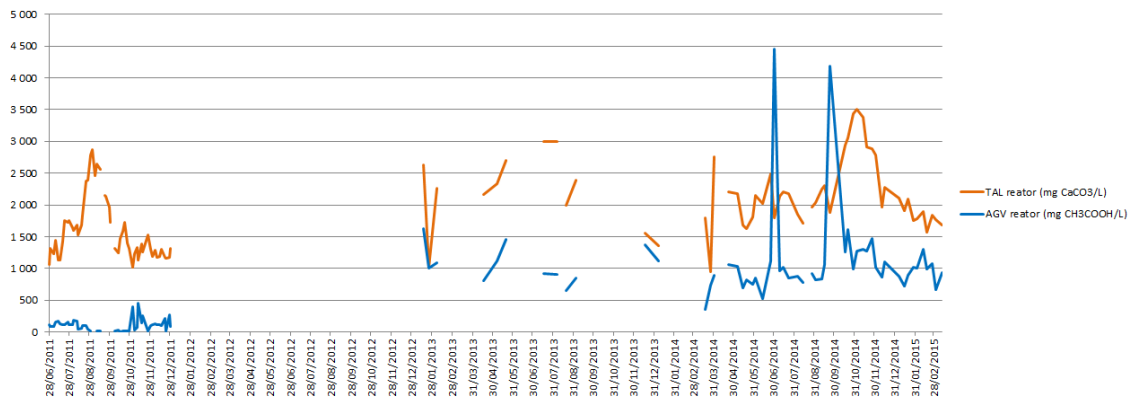


Figura III.6 - Comparação entre os dados das variáveis “TAL reator” e “AGV reator” – ETAR do Seixal

Anexo IV – Resultados do cálculo do coeficiente de correlação de Pearson entre as séries de dados consideradas durante a investigação realizada na ETAR do Seixal no ano de 2014

Tabela IV.1 – Resultados do cálculo do coeficiente de correlação de Pearson (Pinto, 2014)

	Q lamas	pH in	CQO in	LCQO in	MS in	MV in	LMS in	LMV in	CV	MS reator	MV reator	LMV out	LMV removida	pH reator	CQO reator	LCQO out	LCQO removida	TAL reator	AGV reator	AGV/TAL reator	T reator	TRH	TRH (méd. 7 dias)	Q biogás
Q lamas	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
pH in	0,086	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CQO in	-0,229	-0,396	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LCQO in	0,574	0,082	0,537	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MS in	x	x	x	x	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MV in	0,398	0,021	0,406	x	0,931	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LMS in	0,682	-0,006	x	x	x	x	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LMV in	0,586	0,007	x	0,522	x	0,449	0,931	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CV	x	x	x	x	x	x	x	1,000	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MS reator	x	x	x	x	x	x	x	x	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MV reator	0,428	x	0,365	0,155	x	0,227	x	0,093	x	0,910	1	-	-	-	-	-	-	-	-	-	-	-	-	-
LMV out	0,656	x	x	0,559	x	0,215	x	0,496	x	x	0,582	1	-	-	-	-	-	-	-	-	-	-	-	-
LMV removida	0,181	0,158	x	0,073	x	0,560	x	0,683	x	x	-0,296	-0,021	1	-	-	-	-	-	-	-	-	-	-	-
pH reator	0,196	0,228	x	0,151	x	0,104	x	0,184	x	x	-0,329	0,180	-0,147	1	-	-	-	-	-	-	-	-	-	-
CQO reator	0,021	x	0,476	0,096	x	0,429	x	0,070	x	x	0,410	0,086	-0,207	0,108	1	-	-	-	-	-	-	-	-	-
LCQO out	0,665	x	0,058	0,691	x	0,039	x	0,536	x	x	0,150	0,610	0,089	0,092	0,420	1	-	-	-	-	-	-	-	-
LCQO removida	0,084	-0,026	0,687	0,671	x	-0,006	x	0,086	x	x	0,167	0,160	0,006	0,120	0,038	0,158	1	-	-	-	-	-	-	-
TAL reator	0,549	0,369	0,119	-0,025	x	0,142	x	0,074	x	x	0,325	0,007	-0,377	0,271	-0,001	-0,137	0,233	1	-	-	-	-	-	-
AGV reator	0,522	0,725	0,293	0,122	x	-0,051	x	-0,008	x	x	0,410	0,013	-0,424	-0,547	0,141	-0,136	-0,055	0,518	1	-	-	-	-	-
AGV/TAL reator	0,295	0,560	x	x	x	x	x	x	x	x	0,266	x	-0,131	-0,057	-0,005	x	x	-0,035	-0,048	1	-	-	-	-
T reator	0,280	0,325	-0,102	0,174	x	-0,393	x	-0,087	x	x	-0,100	0,051	-0,057	-0,135	-0,153	0,107	0,128	0,421	-0,006	-0,299	1	-	-	-
TRH	-0,900	-0,110	0,318	x	x	0,215	x	x	x	x	0,048	0,167	-0,174	-0,129	0,296	x	x	0,559	0,527	0,026	-0,010	1	-	-
TRH (méd. 7 dias)	0,236	x	x	x	x	0,147	x	x	x	x	0,256	0,240	-0,056	-0,079	0,006	x	x	-0,081	-0,094	-0,094	-0,052	0,254	1	-
Q Biogás	0,442	0,494	0,131	0,221	0,435	0,438	0,445	0,429	0,429	-0,129	-0,120	0,168	0,549	0,529	-0,078	0,107	0,238	-0,223	0,493	-0,031	0,419	0,447	0,209	1

Anexo V – Casos considerados na etapa de previsão das redes neuronais geradas a partir das novas séries de dados existentes na ETAR do Seixal

Tabela V.1 – Casos considerados na etapa de previsão das redes neuronais geradas a partir das novas séries de dados existentes na ETAR do Seixal

	MV in	pH in	MV reator	pH reator	TAL reator	AGV reator	AGV/TAL	Q lamas	TRH	T reator	CV	LMV in	LMV out	LMV removida	Q biogás
17/03/2015	34 050	4,40	14 000	6,12	1 898	1 214	0,64	85,47	23,0	19,35	1,48	2 910	1 197	1 714	476
25/03/2015	11 900	4,90	11 600	6,60	2 300	900	0,39	134,53	14,6	19,12	0,82	1 601	1 561	40	207
07/04/2015	10 850	5,90	5 050	6,14	1 235	686	0,56	38,27	51,3	19,35	0,21	415	193	222	228
16/04/2015	9 250	5,90	9 320	6,74	2 827	574	0,20	31,58	62,2	19,66	0,15	292	294	-2	98
21/04/2015	12 800	5,60	10 000	6,50	2 731	984	0,36	31,36	62,6	19,80	0,20	401	314	88	92
29/04/2015	23 550	6,10	9 000	6,60	2 771	969	0,35	43,11	45,5	19,90	0,52	1 015	388	627	129
05/05/2015	34 600	5,50	10 700	6,40	2 969	1 121	0,38	22,92	85,7	19,99	0,40	793	245	548	85
14/05/2015	23 750	-	11 700	6,50	3 202	1 232	0,38	26,65	73,7	23,59	0,32	633	312	321	118
19/05/2015	25 250	6,20	8 950	6,70	2 591	1 054	0,41	36,89	53,2	21,32	0,47	932	330	601	2
27/05/2015	26 350	5,30	11 700	6,50	3 138	1 021	0,33	12,96	151,5	25,38	0,17	341	152	190	137
02/06/2015	22 950	5,60	12 100	6,50	3 310	650	0,20	63,64	30,8	24,34	0,74	1 460	770	690	5
11/06/2015	4 500	5,80	1 440	6,86	2 561	474	0,19	31,01	63,3	26,94	0,07	140	45	95	74
16/06/2015	21 000	6,00	13 533	6,74	4 049	1 728	0,43	66,18	29,7	23,80	0,71	1 390	896	494	360
24/06/2015	20 600	5,40	15 600	6,47	3 824	1 743	0,46	96,02	20,4	25,05	1,01	1 978	1 498	480	558
07/07/2015	21 100	5,60	15 100	6,50	2 900	1 703	0,59	96,53	20,3	25,73	1,04	2 037	1 458	579	622
16/07/2015	24 250	5,30	16 350	6,50	2 706	855	0,32	104,06	18,9	26,43	1,29	2 523	1 701	822	606
21/07/2015	23 400	5,70	15 000	6,50	2 782	1 071	0,38	97,97	20,0	25,30	1,17	2 292	1 470	823	708
29/07/2015	16 750	5,70	19 000	6,50	2 890	1 061	0,37	112,28	17,5	26,56	0,96	1 881	2 133	-253	783
05/08/2015	19 400	6,50	16 600	7,05	2 877	1 065	0,37	131,67	14,9	26,48	1,30	2 554	2 186	369	436
13/08/2015	31 250	5,80	17 300	6,78	2 956	903	0,31	0,00	-	26,17	-	-	-	-	6
18/08/2015	17 950	5,70	19 150	6,46	2 920	914	0,31	90,31	21,7	26,32	0,83	1 621	1 729	-108	682
26/08/2015	20 350	5,50	17 350	6,28	3 400	938	0,28	135,44	14,5	26,50	1,40	2 756	2 350	406	434

Anexo VI - Localização dos pontos de amostragem e medição das variáveis na ETAR do Sesimbra

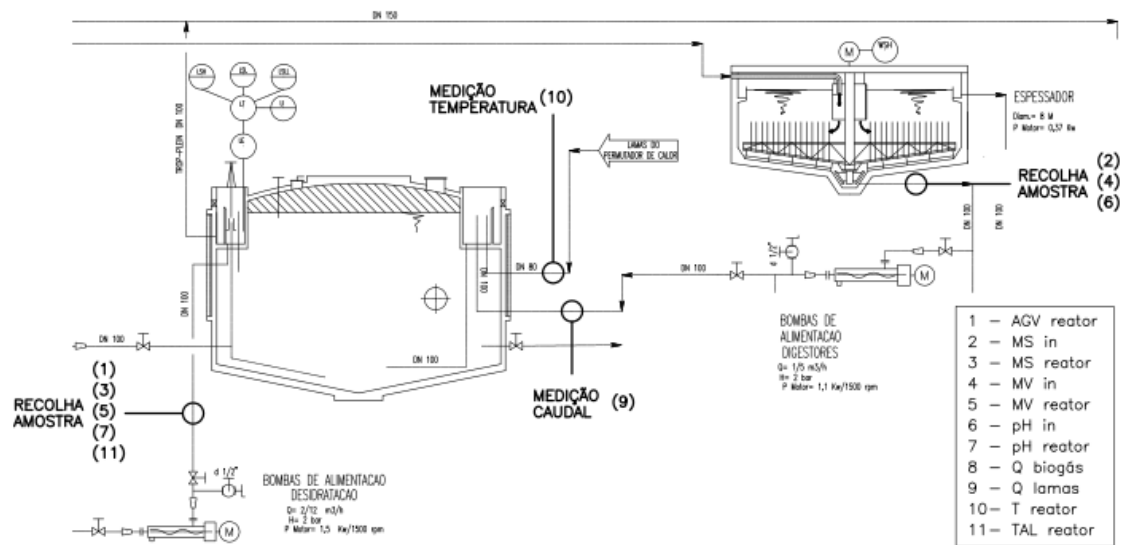


Figura VI.1 – Localização das medições efetuadas e locais de amostragem de alguns parâmetros afetos à digestão anaeróbia – ETAR de Sesimbra

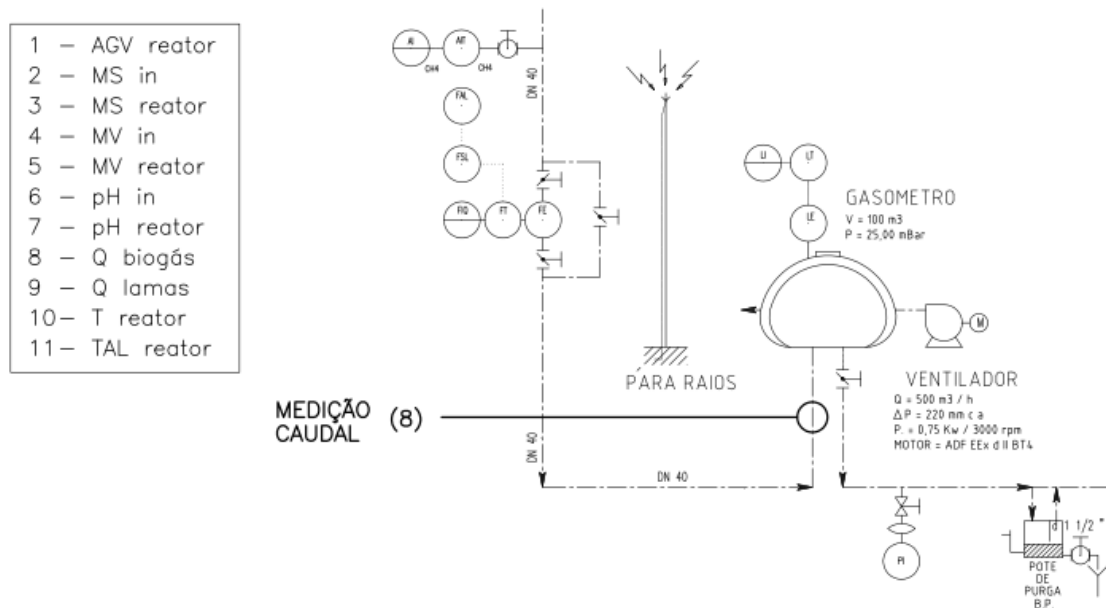


Figura VI.2 - Local onde é efetuada a medição afeta ao caudal de biogás – ETAR de Sesimbra

Anexo VII - Representação gráfica do registo de dados das variáveis utilizadas na investigação realizada na ETAR do Sesimbra

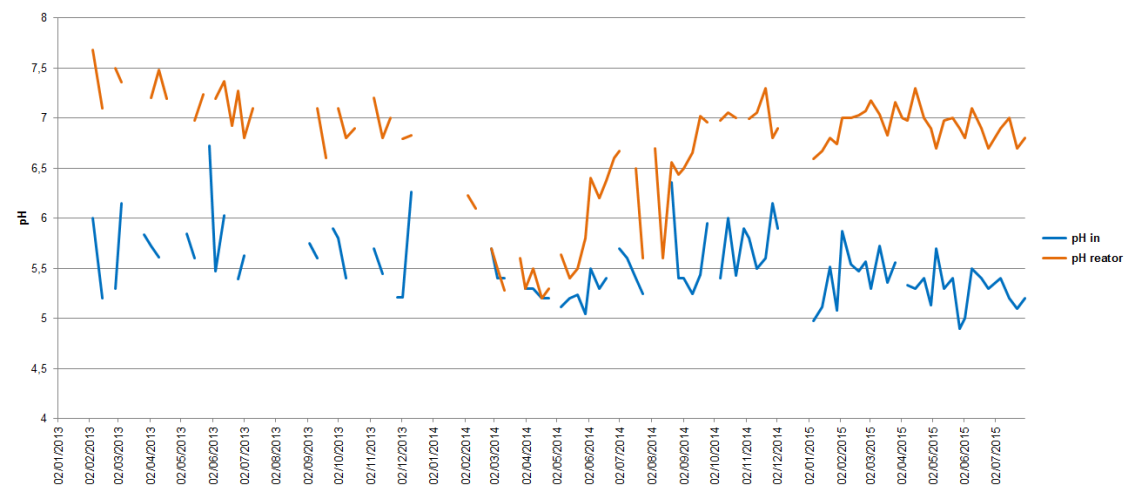


Figura VII.1 – Comparação entre os dados das variáveis “pH in” e “pH reator” – ETAR de Sesimbra

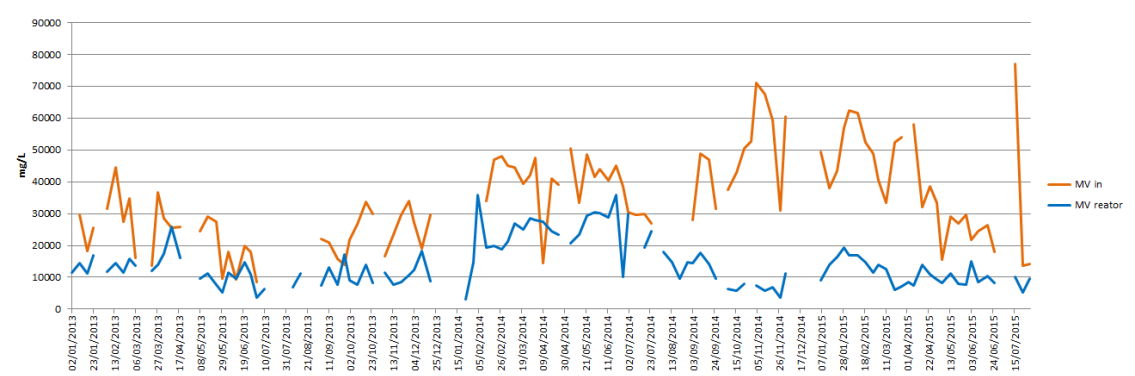


Figura VII.2 – Comparação entre os dados das variáveis “MV in” e “MV reator” – ETAR de Sesimbra

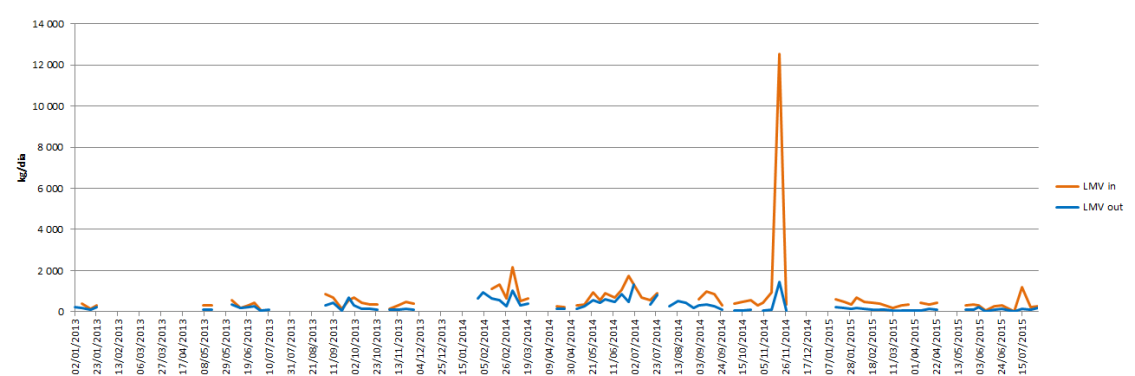


Figura VII.3 – Comparação entre os dados das variáveis “LMV in” e “LMV reator” – ETAR de Sesimbra

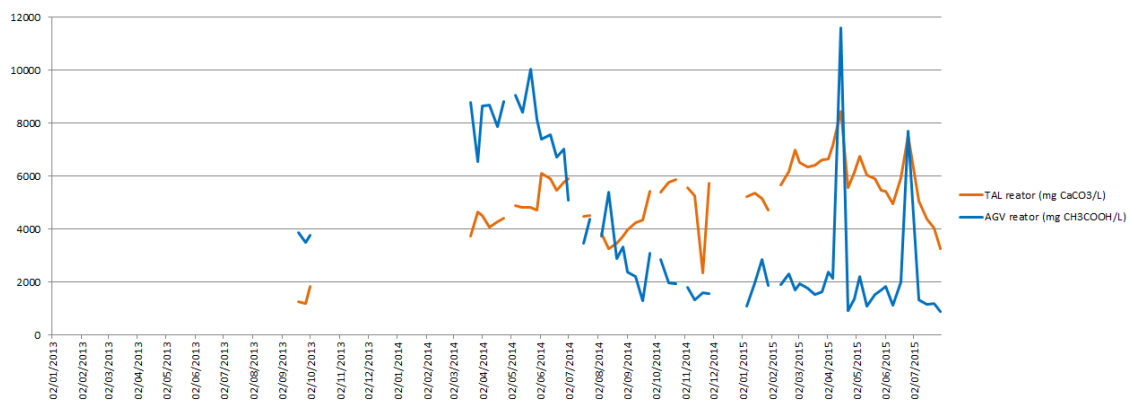


Figura VII.4 – Comparação entre os dados das variáveis “TAL reator” e “AGV reator” – ETAR de Sesimbra

Anexo VIII – Resultados da fase de treino e teste das redes neuronais “2 8H sx” e “3I”, geradas a partir das séries com adição de dados

Tabela VIII.1 – Resultados da fase de treino e teste das redes “2 8H sx” e “3I”

Rede	Variáveis de entrada	Impacto	Nº casos	Data		Resultados	
2 8H sx	AGV reator	36,70%	69	05/02/2013	30/07/2015	Melhor rede	GRNN
	pH reator	31,23%				% más previsões (treino - 30% toler.)	43,64%
	T reator	27,89%				% más previsões (teste - 30% toler.)	71,43%
	TAL reator	4,17%				Erro médio quadrático (treino)	72,31
						Erro médio quadrático (teste)	44,95
3I	T reator	62,46%	41	05/02/2013	30/07/2015	Melhor rede	GRNN
	MV reator	26,13%				% más previsões (treino - 30% toler.)	15,15%
	LMV out	10,96%				% más previsões (teste - 30% toler.)	75,00%
	MV in	0,34%				Erro médio quadrático (treino)	41,76
	AGV/TAL	0,11%				Erro médio quadrático (teste)	69,96
	pH reator	0,00%					